

Responses to Alferov

The Nobel laureate's prizewinning work was accomplished in the Soviet era, but his example and initiative can help Russia rebuild its science and technology. Philanthropists should take note.

It takes a lot of idealism to be a scientist in Russia these days. The first Nobel prize given to a Russian scientist in 22 years cannot mask the fact that the country's research community is facing problems compared to which the needs and complaints of western scientists are mere trifles. Zhores Alferov's prize was won for brilliant work on semiconductors — research that flourished within the Soviet military-industrial system (see page 398). It serves to highlight not only the collapse of that system but also some key opportunities in Russia that, with Alferov's continuing help, must be pursued.

Most of Russian industry has been destroyed rather than restructured, and the state of the science system is pitiful. Although living standards have improved for some, scientists in Russia still live at a subsistence level, the infrastructure and equipment in most institutes are outdated or in need of repair, and funding has only marginally recovered from the collapse of science funding in the 1990s.

Perhaps what gnaws most painfully at Russian scientists' self-esteem is the fact that the profession of science has lost almost all of its former prestige. Science has a long tradition in Russia's major cities — in Moscow and perhaps even more in St Petersburg, "the most abstract and intentional city in the whole round world", as Dostoyevsky described it. Whether under the tsars or under the Soviet rulers, scientists, explorers and inventors enjoyed a high status among Russian intellectuals and the citizens at large.

Ironically, it was one of Russian history's more lucid moments — 'perestroika' (reform) and 'glasnost' (openness) introduced by Soviet President Mikhail Gorbachev — which flipped a strong national science base into bleak depression within a matter of years. Having subsequently embarked on a crash course in raw capitalism, today's Russia is characterized by short-term profiteering and financial speculation, a new type of black economy and a deep divide between rich and poor.

Putin's push

But it would be wrong to dismiss Russia as a wasteland for science. Encouragingly, President Vladimir Putin has not only made science a high priority but is showing more determination in that regard than his predecessor. There are worrying signs of a reversion to centralism, but there is hope that that can be tempered by encouraging the qualities that have helped science flourish elsewhere: individualism and entrepreneurialism. There is scope for science also to benefit significantly from altruism, not only from international foundations and foreign government aid, but also — if history elsewhere is a guide — through the development and encouragement of private philanthropy within Russia itself.

President Putin is encouraging technology incubators and venture funds, so there is a sign that entrepreneurialism can find support. And philanthropy? Private fortunes certainly exist, but the new, moneyed aristocracy shows little insight or interest in the mechanisms of science, innovation, economic growth or social welfare. On the other hand, it was a private foundation that led the way during the worst years of hardship. During 1992–96, the International Science Foundation of the Hungarian billionaire George Soros injected more

than US\$100 million into the states of the former Soviet Union. Isn't it possible that some of Russia's rich could be persuaded to support science? And can President Putin be persuaded to encourage them, and thereby stimulate bottom-up development?

Certainly, Russia cannot, in the foreseeable future, help its science base back onto its feet using public money alone. The Russian leadership must explore all possible ways of encouraging the growth, from virtually zero, of private funding. It needs vigorously to reaffirm the value of civilian science, and it must use economic incentives, such as tax exemptions, to help create private science foundations.

Economic changes

In parallel, it must try to create the economic conditions under which high-technology industry can become re-established in Russia. The Russian parliament's recent vote for a boost in support for the national electronics industry, stimulated by an appeal from Alferov, is a positive step. Central European countries such as Poland, Hungary and the Czech Republic have successfully shown that they are not simply low-wage attractors for multinational companies, but that their once-communist economies can be transformed relatively smoothly and quickly by the right mix of political and private initiative. In particular fields, such as biotechnology, these countries are making significant strides.

All of that being said, Russia needs external help as much as ever. Its traditional affinity for science may not be altogether lost, and it certainly has talented human resources. Moreover, the establishment in 1995 of the Russian Foundation for Basic Research, a self-governing funding agency, has increased scientific competition and quality control. Nonetheless, the country is hindered by the parlous state of most of its public infrastructure and its continual state of crisis management, inflation, corruption and criminality. Therefore, western help is essential. The example of Alferov's laboratory at the Ioffe Physico-Technical Institute in St Petersburg shows how east-west cooperation can lead to successful joint ventures.

Western support of Russian science has little to do with charity, but everything to do with intellectual opportunity. Furthermore, conversion of military science and close cooperation in as many other fields of science as possible are valuable factors in helping maintain international security. NATO, the North Atlantic Treaty Organization, is helping the situation with its science programme, which supports east-west collaborations. And last week's decision to open the European Union's research programmes to Russian scientists (see page 396) is a similarly wise step.

Ultimately, Russia has the human potential to create a new civilian scientific culture that integrates basic research and technology transfer, and which could even come to rely on well-equipped institutes and fair salaries, competition and peer review. The people representing this potential want to be used, but time is moving on. President Putin has responded to Zhores Alferov's achievement by promising further increases in the science budget next year. If Alferov's success can also stimulate private investment, so much the better. ■





On course
International fusion
reactor moves a
step closer
p392



Genetic windfall
Germany answers
the call for post-
genomics research
p393



Nanobacteria?
Doubters ask if
tiny particles are
really alive
p394



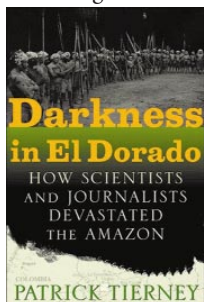
Dambusters
The environmental
costs of dams come
under fire
p395

Anthropologists in turmoil over allegations of misconduct

Rex Dalton, San Francisco

Researchers clashed at the annual meeting of the American Anthropological Association (AAA) over a book that has generated a storm of criticism for anthropologists since it was published earlier this year.

Written by journalist Patrick Tierney, *Darkness in El Dorado: How Scientists and Journalists Devastated the Amazon* describes a study of the Venezuelan Yanomami people. In the book, Tierney claims that data were manipulated, the Yanomami tribe abused, and that research techniques contributed to the deaths of natives through illness or warfare. The AAA has now set up an ethics inquiry to investigate these allegations.



Tierney's book alleges that anthropologist Napoleon Chagnon, from the University of California at Santa Barbara, deceased geneticist James Neel of the University of Michigan and their colleagues may have contributed to a measles or fever epidemic by inoculating Yanomami with a suspect vaccine. The book also alleges that they promoted warfare and violated the culture of the primitive peoples.

Chagnon, who declined an invitation to take part in last week's meeting, has denied the allegations, speaking through supporters and in public statements. Chagnon's own 1968 book, *Yanomamo: The Fierce People*, which describes his studies of the tribe, has since become a major anthropology text.

Tierney's allegations were discussed at two extraordinary sessions of the AAA meeting. The first three-hour session included contributions by a physician who is an authority on measles vaccines, a medical historian, an expert on informed consent, a physician, an anthropologist and tribe member from Venezuela, and William Irons, a Northwestern University anthropologist, who spoke on Chagnon's behalf.

With Tierney sitting at the other end of the table, Irons lambasted the book as rife with falsehoods, accused the author



No apologies: Patrick Tierney defends his book from criticism at the anthropology meeting.

of being duped by Chagnon's enemies and urged AAA members to boycott the book. Tierney's critics also pointed to statements released by researchers at the University of California at Santa Barbara, the University of Michigan and the National Academy of Sciences that have harshly criticized the book (see *Nature* 408, 283; 2000).

After all the panel members had spoken, a visibly shaken Tierney defended his book, asking "everyone to work together" to address the issues raised in it. "Hopefully, we can find some light from this darkness," he concluded. Tierney's talk was greeted with a round of applause that was only exceeded by that given to the Venezuelan tribal member's plea for native peoples.

During the second session the next evening, dozens of anthropologists gave emotional three-minute statements. Several speakers denounced the previous session, saying panellists had shown detestable, colonial "arrogance" as they sought to undermine Tierney's book. Other speakers launched new verbal assaults on Tierney, calling for investigations and censure of some of the anthropologists who had assisted him.

And several speakers lamented that the debate focused too much on the accused

researchers, and not enough on the vulnerable Yanomami. One anthropologist said: "The Yanomami have been used and abused by outsiders for years. What have the Yanomami gotten for 30 years of research?"

Louise Lamphere, an anthropologist at the University of New Mexico and current president of the association, announced that a special panel — drawn from its committees on ethics and human rights — would conduct an inquiry into the allegations. Early next year, the AAA's executive board will consider the panel's recommendations and decide whether to launch a formal ethics investigation.

But the public row over Tierney's book is already prompting consideration of modifications to some of the techniques of anthropology. The AAA's Ethics Committee has begun developing new proposed guidelines for anthropologists conducting field research.

The guidelines will deal directly with many of the issues raised in the book. For example, Tierney alleges that scientists and journalists failed to aid dying Yanomami, and the guidelines will advise researchers on the provision of emergency medical care for study subjects. They will discuss what remuneration subjects should receive (Tierney charges that machete distribution led to more deaths from inter-village warfare), and what constitutes appropriate informed consent, particularly for remote peoples who face language barriers (Tierney alleges that experiments were conducted on Yanomami without their consent).

Scientific leaders in Venezuela and Brazil also announced inquiries. Venezuela placed a moratorium on permits allowing research on indigenous people such as the Yanomami. And the Brazilian anthropological association issued a statement saying it will work with the AAA to strengthen guidelines for studying native populations there.

Anthropologist Jesus Cardozo, who represented Venezuela's office of indigenous affairs at the meeting, says his country is "committed to a fair and impartial analysis" of the allegations of "the most serious violations of human rights". Tierney's book, Cardozo says, "moves us to reflect: were human subjects reduced to human objects?" ■

Partners edge closer to vital step on fusion

Quirin Schiermeier, Munich

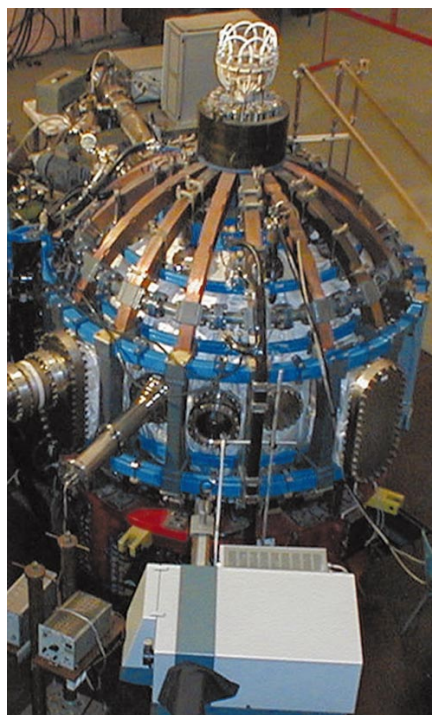
Fusion research received a tentative boost last week when the member states of the European Union (EU) reaffirmed their commitment to build the International Thermonuclear Experimental Reactor (ITER), the experiment regarded as a vital next step towards realizing fusion energy.

Last week, the European Council of Research Ministers gave an official mandate to the European Commission to start negotiations with Russia and Japan over the details of ITER's construction and operation. The negotiations will cover the legalities of ITER operation, the structure of its management, recruitment of staff, their terms of employment and the search for an appropriate site for the reactor.

But the research ministers also asked the commission for a full report on the current state and likely future direction of fusion research, to help them decide whether to proceed with ITER's construction.

The commission must also specify the likely cost of the project, its environmental impact, the timetable for its completion and its place in the EU's Framework research programme.

The design and cost of ITER have changed significantly since the project was first conceived. Under its initial configuration, it would have cost 6.7 billion euros (US\$5.8 billion) and achieved ignition —



Fired up: the international project would build a larger version of this Russian Tokamak reactor.

the point at which fusion becomes self-sustaining.

But after the withdrawal of the United States from the project, which took effect in 1999, it was trimmed down to a version

known as ITER-FEAT (for Fusion Energy Advanced Tokamak). This machine will not attain ignition and will cost an estimated 3.6 billion euros.

Annual European fusion research costs are around 500 million euros, 40% of which come from the EU's research budget, with the rest coming from the member states. The EU share of construction costs for ITER-FEAT must be earmarked in the next Framework programme of research, which starts in 2003.

Fusion scientists hope that the decision will keep Europe on course for participating in ITER's construction. "Last week's agreements are an important step for the European fusion programme, and will take fusion research forward," says Alexander Bradshaw, scientific director of the Max Planck Institute for Plasma Physics in Garching, near Munich, the main base of the European ITER project team.

Despite continuing uncertainty about whether fusion energy is feasible, all EU member states have approved the agreement. But some still have doubts.

For example, Wolf-Michael Catenhusen, the German secretary for research, admits that it was premature for the council to give the go-ahead to negotiate details of the project before the member states had come to a decision about the value of fusion energy in principle.

♦ <http://www.iter.org>

French target research money at allaying BSE fears

Declan Butler, Paris

The French government has announced plans to triple the funding of research on prions, which are thought to cause bovine spongiform encephalopathy (BSE) in cattle, from 2001.

Prime Minister Lionel Jospin announced the increase, from FF70 million to FF210 million (US\$27 million), as part of his government's seven-point plan for dealing with mounting public concern in France about BSE.

The extra money is for research on prion-related diseases, tests to detect the diseases in animals, methods of destroying meat and bone-meal, epidemiological research and the development of therapeutic methods for controlling variant Creutzfeldt-Jakob disease (vCJD), the human form of the disease. The use of meat and bone-meal in animal feed is also being suspended.

The plan will create several new laboratories, and jobs for 120 researchers

and technicians. The many agencies involved in BSE research in France are also to be more closely coordinated.

In the week before Jospin's statement, President Jacques Chirac had made a public statement demanding the immediate banning of meat and bone-meal in animal feed and the systematic testing of all French cattle for BSE. Jospin had preferred to wait for recommendations from government food-safety experts, expected in three or four months, but Chirac's statement forced his hand.

Jospin, quoting the French food safety agency, said at a press conference that there was "no scientific evidence to show that eating meat and milk from cattle represents a health risk".

The French BSE crisis, with 183 cases in cattle and three cases of vCJD, is smaller than that in the United Kingdom. But with the contaminated blood scandal, in which three former government ministers were taken to court for involuntary homicide (see



La vache folle: a French protest against BSE.

Nature 397, 548; 1999), fresh in the public mind and following the Phillips report on the British BSE crisis, both Chirac and Jospin felt obliged to act.

Two of the families of French victims of vCJD are filing lawsuits against "persons unknown" for poisoning and involuntary manslaughter, targeting British, French and European authorities.

Europe boosts genome resource centres...

Alison Abbott, Munich

Brussels bureaucrats can be flexible when it matters. The European Union's (EU) research commission last week announced changes to the Quality of Life component of its fifth Framework programme of research — which is usually thought to be set in stone. The changes mean increased support for genomics resource centres and large-scale genomics research.

The move is a response to complaints from scientists that EU research money is spread too thinly and that the work involved in applying is disproportionate to the low success rate and limited financial returns. They also claim that it fails to support resource centres which they believe are vital to post-genomics research in Europe. Such facilities include the European Bioinformatics Institute (EBI) near Cambridge and the European Mouse Mutant Archive in Rome.

Many EU member states, fearful of long-term commitments, want the Framework programme to fund only research at such facilities, and not their core activities. This recently threw the centres into financial crisis (see *Nature* 405, 723; 2000).

But the commission has now earmarked an extra 25 million euros (US\$21 million) for "genomic and proteomic databases and repositories of suitable animal models". The money is being raised by an internal 1% levy on other parts of the Quality of Life programme.

A rule preventing the support of routine operational costs still holds, as the commission cannot change the legal basis of the Framework programme. But it will be more open-minded in its definition of what can be funded. For example, centres have been informally assured that the expansion of databases and repositories will be eligible.

"There is no question that the extra money is good news," says Graham Cameron, joint head of the EBI. But some researchers are nervous that eligibility depends on how reviewers interpret the rule.

The commission has also shuffled Quality of Life funds to create a pot of 30 million euros for "integrated projects", large initiatives incorporating research, networks and training. A call for ideas in functional genomics for human health has been issued, and the commission will make a preselection before summer.

There will then be a second call for proposals in five chosen research areas, and three integrated projects will be funded. Contracts are unlikely to be signed before February 2002.

The impetus behind these changes comes from two directions. First, the commission wants to show its willingness to support



functional genomics on a large scale, using ideas generated by the research community. Second, it hopes to test out a mechanism it would like to see operate in the sixth Frame-

work programme. This will most probably abandon the funding of individual projects, and will focus instead on large-scale programmes similar to the integrated projects (*Nature* 407, 433; 2000).

The response from the scientific community has been generally positive. "It sounds wonderful," says Hans Lehrach, a director of the Max Planck Institute for Molecular Genetics in Berlin and a spokesman for the German Human Genome Project. "But it seems a slow procedure — in the meantime a post-genomic Celera-2 could have come in and swept away all the rewards."

"Big projects are important," says Joel Bockaert, head of the Centre National de la Recherche Scientifique's Laboratory of Cellular and Molecular Physiology in Montpellier. But, he says, each integrated project would need a research manager, and these are in short supply across Europe.

...as German genomics gets cash windfall

A windfall of money raised by government sales of mobile-phone licences is set to provide a boost for post-genomics research in Germany.

DM350 million (US\$150 million) will be spent over the next three years in building up a 'technology platform' comprising different high-throughput post-genomics technologies, and creating networks of university clinical researchers to use it.

The technology platform will be installed primarily in non-university research centres where genomics infrastructures have already been built up. These include the GSF national research centre in Munich, the Max Delbrück Centre for Molecular Medicine and the Max Planck Institute for Molecular Genetics in Berlin, the Centre for Biotechnological Research in Braunschweig and the German Cancer Centre in Heidelberg.

Decisions about what post-genomics facilities will be provided, and whether some of them should be extended into universities, will be made in consultation with the scientific community over the coming weeks. Possible components include a single-nucleotide polymorphism (SNP) facility, to analyse the small genetic



Research secretary Catenhusen targets technology access.

variations between individuals which may correlate with disease, and a battery of new mouse-phenotyping laboratories, which will 'diagnose', or otherwise characterize, mutant mice.

The profiles of the clinical networks — in categories such as cardiovascular disease, neuropathology or cancer, for example — will also be decided upon within the next couple of months, through consultation with the clinical community, probably followed by a competition.

"Germany will be the first country to take a systematic approach to post-genomics," pledges Wolf-Michael Catenhusen, secretary for research in the federal government. He says that

the approach will give universities, which train the next generation of scientists, access to the newest genomics technologies for their clinical research. Spending could start as early as six months from now.

"This is a great way of bringing clinical researchers on board, with a much wider range of technologies at their disposal," says Martin Hrabě de Angelis, who runs the mouse mutant facility at the GSF. "If someone comes up with an interesting candidate gene for a disease, we can funnel it into the platform and generate a huge amount of information."

Some clinical researchers are unhappy, though, that more of the money is not going to the universities themselves. Clemens Sorg, dean of medicine at the University of Münster and spokesman for the interdisciplinary research groups in clinical medicine, feels that the planned university networks "are to be the servants of the big platform".

But Detlev Ganten, director of the Max Delbrück Centre, denies this will be the case, and says that scientific credit will ultimately go to clinicians. "We will all work to the same end, and the interesting results will come from the clinical networks," he says.

A. A.

Placebos could improve link between medical outlooks

Paul Smaglik, Washington

Study of the placebo effect — the role of medically inactive treatments in healing — could help to bridge the divide between mainstream and alternative medicine, a National Institutes of Health (NIH) meeting was told this week.

Stephen Strauss, director of the National Center for Complementary and Alternative Medicine, has been trying to heal the rift that has, if anything, deepened since Congress forced a reluctant NIH to research alternative medicine in 1992. He is encouraged that 20 other NIH institutes and offices helped to organize the meeting, and that several institute directors took part.

Gerald Fischbach, director of the National Institute of Neurological Disorders and Stroke, described the placebo effect as complex and poorly understood. He suspects that it has some biological basis, perhaps involving specialized nerve cells that respond to the expectation of treatment.

Steven Hyman, director of the National Institute of Mental Health, said that the increasing acceptance of such concepts may explain why researchers are paying more attention to placebos. For example, there is growing evidence that repeated exposure to a stimulus can reconfigure the brain's circuitry.

But it may be difficult to tackle such issues until the study of placebos gains broader acceptance, and psychologists, neuroscientists and social scientists become more willing to join forces. The NIH now offers more interdisciplinary grants, but Hyman is worried that young scientists may shy away from



Sugaring the pill: research may help convince those who find the placebo effect hard to swallow.

the problem, because of the potential for failure in an untested field.

The presence of so many senior NIH researchers at the meeting contrasted with the past hostility towards Strauss's centre. Former NIH director Harold Varmus tried unsuccessfully to cut its budget, even as the overall NIH budget rose. And the institute's last director, Wayne Jonas, was often a lightning rod for criticism. Strauss has received broader support because he appears to be trying evaluate alternative medicine using the best available science, and he sees the placebo effect as key.

"It is a very important issue for our institute because of all the preconceived biases as to how these traditional therapies must be operating — that they must be operating merely through the placebo effect," Strauss says.

David Korn, vice-president for research at the American Association of Medical Colleges, said that Strauss is doing a good job in narrowing the gap between orthodox and alternative physicians. And he applauds Strauss's focus on the placebo effect, which he believes will help all physicians. ■

Researchers fail to find signs of life in 'living' particles

Alison Abbott, Munich

Despite their tiny size, the purported pathogenic microorganisms dubbed 'nanobacteria' are still stirring big controversy in microbiological circles.

A report published in *Proceedings of the National Academy of Sciences* last month (97, 11511–11515; 2000) has failed to confirm biochemical signs of life in the apparently self-replicating particles.

Finnish biochemist Olavi Kajander of the University of Kuopio believes the particles, which are 50–500 nanometres in diameter, are a new type of microorganism. He says that they could be involved in kidney disease. (*Nature* 401, 105; 1999).

The particles all have coats of heavy hydroxyapatite, the mineral that forms teeth and kidney stones. According to Kajander, this coat makes them resistant to attack from environmental assaults — and also to standard microbial and biochemical analysis. But he says the particles appear to be living because microscope analysis reveals broad similarity to bacteria. He also says that they stain for DNA, using a modification of a standard method.

Now some of his experiments have been repeated independently. John Cisar of the US National Institutes of Health in Bethesda, and colleagues from the Food and Drug Administration, have created the particles and reproduced their self-propagating properties.

But the researchers say that the particles' self-propagation and morphology can be explained in terms of microcrystallization, independent of any 'living' mechanism. They could not detect any nucleic acids or proteins in the particles. Moreover, the self-propagation was unaffected by sodium azide, which inhibits respiration. Cisar says that "there is a need for hard molecular evidence" to support a claim for a biological base.

Kajander and his colleagues, including some clinical researchers, counter that the special properties of nanobacteria mean that standard biochemical analyses are not necessarily applicable. "We have evidence that the particles are living," insists Neva Çiftçioğlu, also at Kuopio. "We are not fanatics, we are scientists."

And Franklin Cockerill, chair of microbiology at the Mayo Clinic, Rochester, says he will continue trying to extract DNA from the nanobacteria. He hopes to sequence the genome for the particles, which, he suggests, could indicate the smallest number of genes required for life. ■

Biotech report dubbed 'worthless'

Haim Watzman, Jerusalem

Research directors at Israel's universities say that an upcoming government-sponsored report on biotechnology policy will be worthless because they were not consulted by its authors.

"The authors of the report are completely unaware of the reorganization the universities have undergone to take into account the needs of high-tech," says Yair Aharonowitz, vice-president and dean for research at Tel Aviv University.

Aharonowitz and colleagues at five other universities sent a letter of protest to Carmel Vernia, chief scientist at the Ministry of Industry and Trade. Earlier this year Vernia commissioned a report on the state of Israel's biotechnology industry from the Monitor

Group, a consulting firm. Although Monitor has until the end of the year to submit its report, the Israeli press has reported that it will call for the diversion of money from basic research to work with clearer commercial potential.

Any such diversion would be a "gross error", Menachem Magidor, president of the Hebrew University of Jerusalem, declared in a recent interview. Aharonowitz says that Monitor got all its information from industry, and made no contact with five of the six Israeli universities with life-science faculties.

Vernia declined to comment, but a spokesman said that he had "recommended that Monitor meet with the research vice-president of the Ministry of Industry and Trade of the process of writing its report". ■

Governments urged to rethink dam projects

Mark Schrope

The environmental and social cost of major dam projects around the world has been unacceptably high, even though many of them have provided significant economic benefits, says a global assessment that is expected to influence future consideration of dam developments.

The World Commission on Dams, which released its report in London last week, proposes new frameworks for deciding whether dam projects are really needed and for planning those that do proceed.

The report has been welcomed by ecologists and health researchers, many of whom have criticized national governments and agencies such as the World Bank for ploughing ahead with large dam projects in the face of environmental objections. "It represents a unique body of knowledge that has not been accessible before," says Adrian Sleight, a researcher into tropical public health at the University of Queensland in Brisbane.

The commission is an international 12-member panel representing such diverse interests as commercial dam builders, indigenous peoples and environmental organizations. It functions independently, but is sponsored by the World Bank and other interested organizations.



It estimates that the 45,000 large dams around the globe have cost over \$2 trillion to build and are responsible for about 20% of the world's power and 15% of its food supply. The report says that although dams can have positive ecological effects, such as establishing new wetland habitats, their overall effect on the environment has been negative because of loss of habitat, species reductions and other problems. In addition, dams have displaced between 40 million and 80 million people from their homes.

Deborah Moore, a member of the commission and a consulting scientist with Environmental Defense, a non-profit organization based in New York, says that few of the dams have been properly evaluated on their performance in producing energy, controlling floods or delivering water supplies.

Damning verdict: the World Commission on Dams has found that the social and environmental costs of projects such as Pakistan's Tarbela dam, shown here, can sometimes outweigh their economic benefits.

But the commission's study suggests that large dams have typically not met their intended goals.

The report calls for long-term monitoring of dam performance to maximize efficiency and reduce harmful effects. And it strongly urges planners to consider potential alternatives to dams, such as better management of existing resources and interconnection of power grids.

The commission adds that when dam projects do proceed, social and environmental effects should be considered just as important as economic ones — in contrast to current standard practice. It also urges that people affected by projects be allowed to participate in planning, and that developers should be held contractually accountable for accomplishing agreed goals. ■

CHRISTINE OSBORNE/CORBIS

Aventis gets short shrift over release of modified corn

Jessa Netting, Washington

The US government looks unlikely to bow to demands from Aventis that it temporarily approve the company's genetically modified StarLink corn for human consumption, following the inadvertent and embarrassing release of the strain into the food chain.

The US Environmental Protection Agency (EPA) appears unimpressed by data submitted in support of the request by Aventis, the North Carolina-based manufacturer of the corn. The EPA's preliminary evaluation questioned the company's interpretation of studies on the potential for allergic reactions to the corn.

The EPA made its assessment in preparation for a public meeting near Washington on 28 November at which it will solicit comment from the public and from a panel of scientists. The regulatory agency will receive the panel's final recommendations on 1 December.

A coalition of US environmental groups concerned about genetically engineered foods initiated the furore in mid-September when their tests found traces of StarLink DNA in taco shells. The discovery led to a massive recall of more than 300 food brands.

The submission of new data is part of a petition by Aventis to allow StarLink corn to be present temporarily in processed foods, to avoid more recalls of products already on the market, says a company representative.

Debate over this strain of corn has been heated. The strain, approved in 1998 for use in animal feeds, is engineered to produce an insecticidal protein related to others already in widespread use. But the protein, Cry9C, does not break down at certain

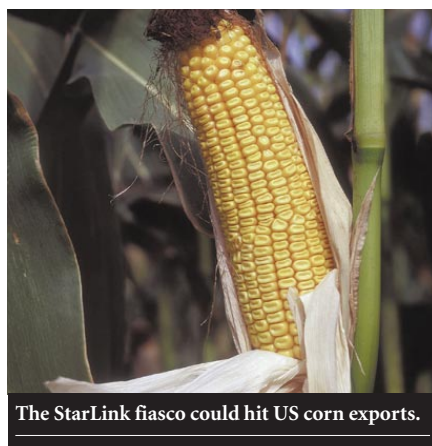
temperatures and conditions, which has caused alarm about its potential to trigger human allergies.

Despite these concerns, an Aventis spokesman says the company is convinced that StarLink will not cause allergic reactions. Another Aventis representative, Rhonda Barnat, adds: "This is a technical regulatory issue, not a product safety issue."

The EPA has criticized the company's analysis of the new data. EPA official Stephen Johnson says: "We don't think they have provided us with sufficient information [on the allergy issue]." The EPA also finds fault with the way Aventis determines allergic potential and uses peanut protein, a potent allergen, as a measure of safety.

"It is clear that the company violated its licence and that is an outrage to us," says Johnson, adding that the agency needs to review the new information further before making any final decisions.

Taiwan also reports finding StarLink in corn grits, says Tim Galvin, an official at the US Department of Agriculture's Foreign Agricultural Service. And Japan has found the variety, which it had not approved, in 10 of 15 samples of a baking product. ■



The StarLink fiasco could hit US corn exports.

PHOTOBAR.COM

Computer billionaire sets up environment and health foundation

Washington The co-founder of the computer chip manufacturer Intel and his wife last week established the multibillion-dollar Gordon E. and Betty I. Moore Foundation to fund research and education on health and the environment.

The foundation — the latest of several huge acts of philanthropy by US computer billionaires — plans to back projects that would not win government support.

“We want to find good opportunities which, although they may be a bit riskier, offer the possibility of higher returns,” says Lewis Coleman, a bank chairman appointed as the foundation’s president. The San Francisco-based foundation’s first act will be to hire staff and set up grant procedures.

Second fruitfly sequence waits in the wings

Washington The fruitfly *Drosophila melanogaster* is one of the select band of species to have had their genomes fully sequenced. Now it could soon be joined by a second member of the genus, *D. pseudoobscura* (see page 400).

The *D. melanogaster* sequencing project — a collaboration between the *Drosophila* Genome Center at the University of California, Berkeley, and Celera of Rockville, Maryland — proved so efficient that US\$14 million of the allocated funding remained unspent. A second species could be sequenced for \$4 million, says Gerald Rubin, who heads the Berkeley centre.

This would help to make sense of the *D. melanogaster* genome — just as the mouse genome is expected to aid annotation of the human sequence. Rubin has recommended *D. pseudoobscura* because it has a relatively small genome and is neither too closely related to *D. melanogaster* nor too distant from it.

The US National Institutes of Health is not expected to solicit a call for proposals for the project until its budget for the coming year has been agreed. Nonetheless, Rubin says he is “cautiously optimistic that we will have the sequence within a year”.

Australian takes over at Biosphere 2

Sydney Barry Osmond, an Australian plant biologist, has been named president of the Biosphere 2 Center, the controlled-environment facility in Arizona run by New York’s Columbia University.

Osmond was formerly head of biological research at the Australian National University in Canberra. He says the US\$200 million centre, built by Texan oil magnate Ed Bass, is uniquely equipped to do much-needed research on ecological systems.

Osmond adds that he has become disillusioned with what he sees as the long-term damage done to Australian science by five years of cuts to research and the universities by Prime Minister John Howard’s conservative government.

EU forges links with Russia for joint research

Brussels The European Commission has signed a science and technology cooperation agreement that will allow Russian scientists to participate in all non-nuclear European Union (EU) research programmes.

EU research organizations will, in turn, be allowed to participate in Russian research programmes — although these have been massively cut back in the past decade. Russia spent 3% of its gross domestic product on research in the 1980s, compared with just 0.3% today.

The new accord is part of a more general ‘Agreement on Partnership and Cooperation’ between Russia and the EU, to promote

closer political relations and improve economic collaboration.

Indian telescope begins operation

New Delhi India's new optical telescope atop Mount Saraswati in the Himalayas has begun operation, after 'first light' on 27 September. But officials say it may take a few months to recruit staff and iron out the logistics of operating the world's highest observatory, 4,500 metres above sea level.

The observatory, which cost 380 million rupees (US\$8 million) and houses a 2-metre-diameter telescope, is one of several new astronomy facilities being built by the



Viewing highlight: the new telescope at Hanle is situated 4,500 metres above sea level.

Bangalore-based Indian Institute of Astrophysics. Its site at Hanle, although a 10-hour mountainous drive from the nearest major town, was chosen after a survey of viewing conditions for visible, infrared and sub-millimetre wavelengths.

Spanish scientists complain of 'underdevelopment'

Barcelona Spanish scientists have drawn up a manifesto condemning what they call their country's "scientific underdevelopment, in spite of the extraordinary political, economical and social advances since the arrival of democracy".

The Manifesto for a Social Pact for Science and Technology was delivered to science minister Anna Birulés — who endorses it — on 13 November, when scientist Margarita Salas was presented with the Universal Spaniard award from the prestigious Fundación Independiente.

The manifesto states that, for cultural and historical reasons, science in Spain "suffers from a series of chronic maladies", and points out that public investment in non-military research has actually fallen since 1990. It seeks a commitment by the state and society to "solve in the medium term the gap between current and necessary resources".

Blair condemns 'anti-science' protesters

London British Prime Minister Tony Blair has pledged to defend scientists and biotechnology companies against protesters intent on obstructing their work. He told the European Bioscience Conference, held in London last week, that his government "will not tolerate blackmail", warning that "anti-science" attitudes could undermine medical and economic progress.

Blair acknowledged that biotechnology presents difficult ethical questions and concerns over safety. But he argued that scientists must be allowed to conduct the research needed to judge the issues. "There are legitimate concerns," said Blair. "But to make heroes of people who are preventing basic scientific research taking place is wrong. It is to substitute aggression for argument."

The British government has become increasingly worried about the influence of protesters opposed to field trials investigating the environmental effects of genetically modified crops who have cut down many of the plots before the experiments were completed. Ministers are also worried about the intimidatory tactics of animal-rights extremists who have targeted facilities and individuals involved in animal experimentation.

Russia's prize fighter

When Zhores Alferov won a share of this year's Nobel Prize for Physics, he restored pride to Russian science. But can he exploit his celebrity status to move research up the political agenda? Quirin Schiermeier investigates.

On 10 October this year, Zhores Alferov received a phone call from Stockholm that propelled him out of the lab and into the headlines. But it was the phone call that followed which could prove the key to boosting the prospects of Russian science.

The first call made Alferov a Nobel physics laureate, in recognition of his pioneering work in developing semiconductor heterostructures, now widely used in the microelectronics industry. The other came from the Russian president, Vladimir Putin, offering his congratulations and paving the way for Alferov's increasing influence on Russian science policy.

A few days after the announcement,

at a confidential meeting, Putin accepted Alferov's suggestion to set up an advisory council of science and technology experts. Although the full significance of this move remains unclear, Russian researchers are monitoring keenly Alferov's emerging status as Putin's unofficial science adviser.

Could it be that a Nobel prize — Russia's first in science since 1978, when Pyotr Kapitsa won for his work on low-temperature physics — will help restore science and technology to the centre of political thinking in Moscow? If so, it will be fitting that Alferov was at the centre of events. For his story, and that of the centre he directs, the Ioffe Physico-Technical Institute in St Petersburg, exemplifies the roller-coaster ride that Russian science has experienced.

Scientists of status

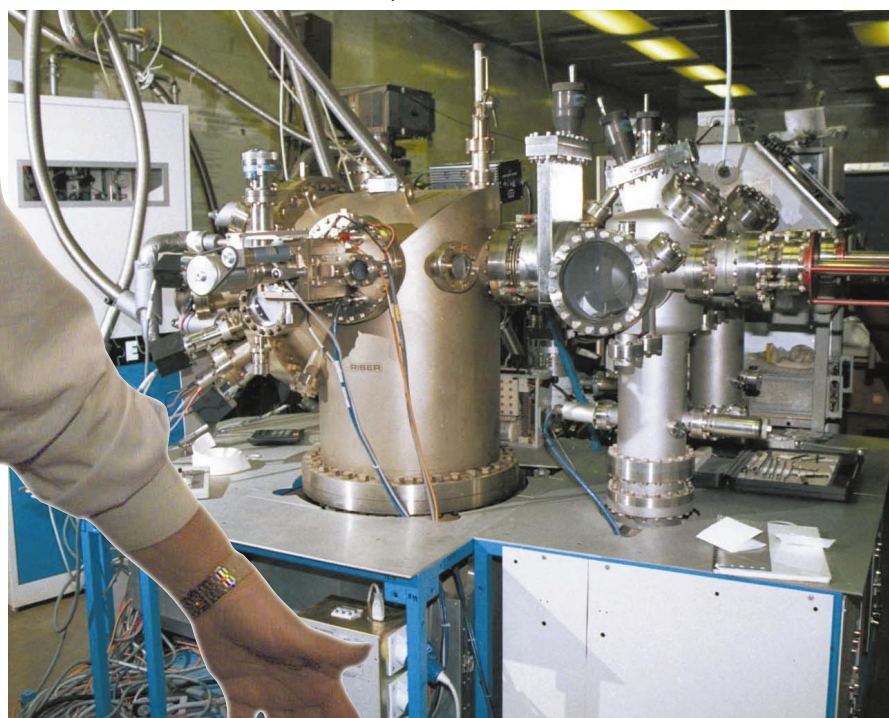
Researchers such as Alferov became a privileged élite after the Second World War, only to see their empires crash when the Soviet Union collapsed. Today, his is just one of many Russian labs that depend largely on western agencies and companies for their financial survival.

A proud and eloquent man who is still loyal to Soviet traditions — he is a Communist member of the Duma, or state parliament — Alferov's western contacts and his reputation for scientific excellence have ensured the Ioffe institute's survival. "I hope the Nobel prize will help us further to recover before it is too late," he says.

The institute was founded in 1918 by Abram Ioffe, a student of Nobel laureate Wilhelm Conrad Röntgen — who won the first physics Nobel in 1901 for his discovery of X-rays. It quickly became a premier centre for physics. But the strongest growth came during the postwar competition with the United States for global superpower status.

Alferov chooses not to emphasize the Cold War factor. "The Soviet leadership was clever enough to support science and scientific education in general as an important branch of human activity," he says. But much of this investment was designed to underpin the powerful military-industrial complex. Alferov's prizewinning work, conducted at the Ioffe institute in the 1960s and 1970s, benefited from the largesse. And by the time he took over as director in 1987,

Man and machine: Zhores Alferov and the molecular beam epitaxy equipment used to make semiconductor heterostructures — these layered structures have revolutionized microelectronics.



the institute's annual budget had grown to US\$80 million.

But the days of plenty ended abruptly with the collapse of the Soviet Union in 1991. As power passed to Boris Yeltsin's Russian Federation, science spending went into free fall. By 1992, the Ioffe institute's state funding had plummeted to around \$4 million per year. Survival-level salaries ate up this budget — buying new equipment was out of the question.

Survival skills

So what kept the institute going? "We would not have made it alone," says Andrei Zabrodskii, a solid-state physicist and vice-director of the institute. Foreign help was very important, with Hungarian-born billionaire and currency speculator George Soros leading the way. Between 1994 and 1996, scientists at the Ioffe institute received 80 research grants worth a total of \$2 million from his International Science Foundation.

Sources of foreign funding have since diversified. Today, about a third of the institute's annual budget of around \$8 million comes from abroad. Here, Alferov's international reputation was invaluable. At the height of the Cold War, he was one of the trusted élite of young scientists who were given permission to visit the West. In 1970, Alferov spent six months in the lab of Nick Holonyak at the University of Illinois at Urbana-Champaign. There he carried out groundbreaking work on the structure and properties of semiconductor lasers.

Indeed, Alferov's first major award, the 1971 Stuart Ballantine Gold Medal of the Franklin Institute, came from the United States. Already a world leader in his field, Alferov was in hot demand to speak at international conferences. But unrestricted travel was impossible, even for someone of Alferov's growing status, and so the Ballantine medal had to be sent to him by mail.

Nevertheless, Alferov prospered within the Soviet system — the spaces between the bookcases in his vast, wood-panelled office are adorned with awards and certificates, and a conspicuous red flag. Visiting his institute, it is clear that junior colleagues regard him with a respect that approaches awe. He is engaging and charismatic, expansive in his gestures and quick to laugh. And although the Ioffe institute's current diminished circumstances must be an immense source of frustration, Alferov does not let it show. "Despite all our difficulties," he says, "the Ioffe institute is still home to some high-quality research, particularly in plasma physics, astrophysics and semiconductor physics."

Much of this research is financed by foreign grants. Projects in nanotechnology, for example, are supported by INTAS, an independent body formed by the European Union to promote East–West scientific coop-



Past glories: the Ioffe institute established its reputation as a premier centre for physics.

eration. Similar support comes from the NATO Science Programme, and the Civilian Research and Development Foundation — a charity based in Arlington, Virginia, partly funded by the US government.

Another important source of funds is the International Science and Technology Center (ISTC), established in 1992 by the European Union, Japan, Russia and the United States. The ISTC redirects activities formerly related to weapons research into civilian projects. The Ioffe institute's spherical Tokamak was completed in 1998 thanks to a \$1 million grant from the ISTC. It is used to investigate the physical properties of spherical plasmas — aimed at decreasing the costs of fusion reactors.

But Germany is the Ioffe institute's most important international partner. At any one time, more than 25 Ioffe staff are working in Germany, many of them supported by the Alexander von Humboldt Foundation. Other support comes from the private Volkswagen Foundation, which, since 1993, has backed Alferov's research with a total of \$250,000. Furthermore, the Ioffe institute is the only non-German institution affiliated to the German science ministry's NanOptics programme, which investigates the use of nanostructures in lasers and other optical devices.

Such grants — and support from foreign companies — have ensured that Alferov's own research area has remained buoyant. The institute's division of nanoheterostructure physics has contracts with companies in South Korea, Germany and China. For Alferov, the great disappointment is that his work is not benefiting a home-grown microelectronics industry. "The inability to transfer our results is the country's main problem," he says.

In 1985, Russia's electronics industry was the world's third largest, after those of the United States and Japan. But with little emphasis on consumer electronics, the industry was decimated by the collapse of the military-industrial complex. "It was destroyed,"

says Alferov, who hopes to use his influence to reverse the damage. On his initiative, the Duma earlier this month made a start by voting an extra \$16 million into the 2001 budget to develop electronic technologies.

Faith in the future

Alferov believes Putin and the new science minister, Alexander Dondukov, will put more emphasis on research. They have already promised a 10% increase in funding, and deploying research to help rebuild Russian industry seems to be a high priority.

But in the long run, the future of Russian science might depend on its ability to exploit its human potential. In that regard, the Ioffe institute's educational centre, established in the 1930s, could provide a model. Ioffe's kindergarten, as it is known, takes talented secondary school students and tries to turn them into the researchers of tomorrow. Around a quarter of its intake stay in science, and the best graduates join the institute's staff. Thanks to this, the Ioffe institute still boasts a scientific staff of some 1,300. The school has already benefited from Alferov's previous political connections: in 1998, it moved into a new building near the institute, financed in large part by an \$8 million donation from former prime minister Viktor Chernomyrdin.

In Russia, which has long suffered from a separation between the educational system and research — most of which is conducted in the institutes of the Academy of Sciences — the centre is highly unusual. Alferov believes such initiatives are vitally important, and he intends to spend a significant proportion of his Nobel prize money on the centre. "We may have an abundance of problems," he says, "but we certainly have no lack of scientific talents."

Quirin Schiermeier is *Nature's* German correspondent.

♦ <http://www.nobel.se/physics/laureates/2000/public.html>

♦ <http://www.ioffe.rssi.ru>

Fruitfly centre spreads its wings

The world's main supplier of exotic *Drosophila* species has had a poor record for customer service. But, as Rex Dalton found, it is having a facelift that should boost studies of evolutionary genetics.

Six years ago, evolutionary biologist Greg Spicer of San Francisco State University was gearing up for a comparative study of developmental genetics. He needed stocks of 50 different species of the fruitfly *Drosophila*, and placed his order with the National *Drosophila* Species Stock Center, then based at Bowling Green State University in Ohio. 'Sorry,' replied the centre's operators — there was a strict limit of ten species per request. So Spicer had to concoct a Machiavellian subterfuge using overnight express mail services and colleagues at other universities to acquire the other 40 species.

Then, when the vials of flies began arriving, Spicer discovered another problem: as many as three species in some batches of ten had been identified incorrectly — forcing him to go through the whole charade again. "The centre was awful," complains Spicer, one of several fruitfly biologists who found that its standards of customer service left much to be desired. A leading *Drosophila* geneticist adds: "It is a unique resource, but it hasn't been very well curated over the past few years."

But fruitfly researchers, and officials with the US National Science Foundation (NSF), which funds the centre, hope that such problems are a thing of the past. In October, the centre's colonies started moving to the University of Arizona at Tucson. There, the centre will be operated by Therese Markow, professor of ecology and evolutionary biology and director of the university's Center for Insect Science. "We plan to expand services considerably," she promises. Limits to the number of species that scientists can order have been scrapped. And Markow's team is working to purify the colonies, to ensure that researchers receive the exact species they order.

Drosophila has long been one of the most important model organisms for studying genetics and development. Much of this work involves one species: *D. melanogaster*. But the genus contains close to 2,000 species, and this enormous diversity also makes it ideal for comparative and evolutionary studies. Different *Drosophila* species live in different climates and eat different things; some are able to exploit unusual food sources because they can detoxify noxious chemicals. Decades of genetic studies on *D. melanogaster*, plus the publication of its complete genome sequence earlier this year¹,

have laid a solid foundation for making comparisons with and between other members of the genus to uncover the genetic basis of particular adaptive traits. "If you know the genetics of one species then you can ask about genetic differences in the others," says Markow. "You can also experiment with these animals in ways you can't with larger animals."

As well as studies that seek to identify the genetic changes that lie behind the evolution of specific adaptations, comparisons between *Drosophila* species can provide a route into the more fundamental territory of evolutionary genetics. Active areas of inquiry include the evolution of sex chromosomes² and the organization of the genome³. Comparative studies of *Drosophila* have also been instrumental in casting doubt on the widely quoted theory that sexual reproduction evolved primarily to help purge the genome of harmful mutations⁴. "*Drosophila* offers some wonderful opportunities," says Brian Charlesworth, an evolutionary geneticist at the University of Edinburgh.

A thankless task

The National *Drosophila* Species Stock Center should be an invaluable resource for evolutionary genetic studies. It maintains colonies of some 1,400 strains of fruitfly representing 265 species. And although it is billed as a national centre, it serves researchers worldwide — there is no comparable facility outside the United States. But the centre's poor performance in recent years has frustrated many of the scientists it was established to serve.

Fly collections for the stock centre began in the 1940s at the University of Texas at

Austin. Scientists searched the world eagerly for new *Drosophila* species in deserts, forests and grasslands. In 1982, the centre moved to Bowling Green, where it was run by geneticist Jong Yoon and his wife Kay — who declined to be interviewed for this article. Their job was not easy: some fruitflies have exotic dietary requirements, others are prone to disease, and species identification is a constant problem — exacerbated by the ease with which colonies become contaminated by other species. And, as anyone who has worked in a *Drosophila* lab will tell you, maintaining breeding colonies of flies is smelly, tedious work.

The Yoons' task was made harder by systematics' fall from favour in the 1980s, as universities channelled their resources into molecular biology. The number of specialists who could identify *Drosophila* species — a difficult task involving microscopic examination of the flies' genitalia — dwindled. "No one knows how to identify *Drosophila* any more," laments David Grimaldi, a curator of entomology at the American Museum of Natural History in New York. "It is amazing. I do identification of *Drosophila* for scientists from everywhere." Even the US Department of Agriculture, which employs legions of entomologists, must from time to time call on Grimaldi's expertise.

Tough decisions

Although they acknowledge these difficulties, Spicer and others say that the centre's performance was unacceptable. And in 1991, it was reviewed by a team of independent scientists brought in by the NSF — in response to concerns that the centre "was not serving the community as well as it might", according to NSF records. Among the problems identified by the review team was that, although Jong Yoon had implied that the centre had a functioning advisory panel, in reality it had not "met or functioned ... in a meaningful way for years".

Identifying *Drosophila* species is tricky, and requires examination of the flies' genitalia under a microscope.



The report added: "It isn't nice to try to pull the wool over the reviewers' eyes."

Despite such concerns, the review team gave the stock centre a rating of "very good". The scientists involved now acknowledge that many of the centre's users will have greeted this assessment with disbelief. But review-team member Kathleen Matthews, a geneticist who heads the NSF-funded *Drosophila* Stock Center at the University of Indiana at Bloomington, which maintains about 6,400 different genetic stocks of *D. melanogaster*, says that the reviewers faced a dilemma. If they were too harsh, the facility might have lost its NSF funding. As no one was pushing to take over the centre, there were concerns that the collection would then be lost to science. "We decided to recommend changes to modernize and make the centre function better," says Matthews.

But despite some improvements, problems persisted — until, in 1999, the NSF decided to find the centre a new operator. Even then, Arizona was the only bidder. The NSF will provide Arizona with some \$1 million over the next five years to set up and operate the facility.

But the fact that problems persisted in Bowling Green for so long, and the circumstances of the 1991 review, raise questions about the NSF's funding and appraisal of such centralized facilities. Gerald Selzer, the NSF official responsible for the agency's 17 stock collections of research resources, says that the difficulties have revealed the limitations of the competitive funding process in cases in which academics are not queuing up to run an important facility. The only alternative would be for the NSF to run such facilities as federal labs, for which there is no precedent.

Bright future

Fruitfly researchers are now looking optimistically to the future. One of Markow's first moves will be to establish a research programme in comparative functional genomics. "The future of the centre is in genomics. It is going to be unbelievable for that type of research," enthuses Spicer, who has collaborated with Markow on comparative studies of *Drosophila* reproductive organs^{5,6}. Genomic technologies, such as the use of DNA microarrays to chart differences in gene expression, will transform such studies, he says. Markow's thrust into functional genomics has also been boosted by the expectation that the US National Institutes of Health will support the sequencing of the genome of a second *Drosophila* species: probably *D. pseudoobscura* (see page 396).

In the spring, Markow intends to host a workshop on *Drosophila* identification and husbandry, which she hopes will begin to revive the skills needed to curate fruitfly stocks. A new collections manager is also



Gotcha: Markow searches the Arizona desert for additions to her stocks of *Drosophila*.

being trained to operate the stock centre. In the meantime, given the dearth of fly systematists, Markow has enlisted a retired Arizona professor, Bill Heed, to help catalogue the collection. Any contamination of stocks will mean that some species have to be collected again from the wild. But Markow hopes not to stop there, and has plans to add more species to the collection in the near future.

Once the collection has been validated, stocks will be available for ordering over the Internet — another service that was never implemented at Bowling Green. That change is indicative of Markow's new broom, say the facility's users. "The centre is going into amazingly good hands and should be a tremendous resource," says Scott Pitnick,



an entomologist at Syracuse University in New York.

Rex Dalton is Nature's West Coast US correspondent.

1. Adams, M. D. *et al. Science* **287**, 2185–2195 (2000).
2. Bachtrog, D. & Charlesworth, B. *Curr. Biol.* **10**, 1025–1031 (2000).
3. González, J. *et al. Chromosom. Res.* **8**, 375–385 (2000).
4. Keightley, P. D. & Eyre-Walker, A. *Science* **290**, 331–333 (2000).
5. Pitnick, S. *et al. Evolution* **51**, 833–845 (1997).
6. Pitnick, S. *et al. Evolution* **53**, 1804–1822 (1999).

Revolution in references: give readers a chance by putting page numbers

Sir — In *On the Origin of Species*, Charles Darwin observed that Malthus's doctrine applied to the animal and vegetable kingdoms. I am a lawyer who recently published a book that required me to analyse hundreds of scientific books (including Darwin's *Origin*), book chapters and articles. In nearly all, a citation to this fact would read this way: "Darwin, C. (1859). *On the Origin of Species*. London: John Murray". Unless readers sleuthed through the entire 490-page book, they would have had to take my word for what Darwin said. The problem was nearly as bad for citations to articles. Standard scientific notation does not appear to require page references for citations, though sometimes quotations get them.

I was surprised by this practice. Standard legal writing requires authors to provide the pages on which they rely for a proposition. It doesn't matter whether the source is a two-page letter or a thousand-page book. That way readers can easily locate the proposition cited.

I have spent hours, even days, sifting scientific articles and books for referenced material. Once I found a citation to a proposition so unusual and pertinent that I purchased the book and read the entire thing. When I failed to locate the cited proposition, I read the book again. Still unable to find it, I tracked down the author and begged her to give me the page numbers. With apologies, she admitted that what I sought was not in the book.

Before me is a book chapter I have written, to be published by a scientific press. When I turned it in to the publisher, I gave footnotes citing the pages at which every proposition upon which I relied could be found. I was informed that this is not good scientific notation. The publisher returned the chapter with instructions for me to go through each footnote and delete the offending references to the exact pages.

What is going on? A physicist friend says that some colleagues aim to outline their achievements while giving away as little as possible to competitors. A primatologist friend believes that scientific specialization is now so common that scientists write only for colleagues in their disciplines, who can be assumed to have read everything that the author has.

Perhaps scientists are simply too busy and trusting to cross-check their colleague's claims. But as a legal outsider peering into the scientific world, I appeal to the instinct of the scientist for precision and accuracy. Require, or at least permit, page references

to cited material. And stop driving me crazy. (NB: the Malthus observation is found on page 63 of the first edition.)

Steve Wise

Wise & Slater-Wise, P.C., 896 Beacon Street, Suite 303, Boston, Massachusetts 02215, USA

The Editor replies—*Nature* now requires authors citing a section of a stand-alone text (for example a book or thesis) to provide the number(s) of the cited pages. Authors need not provide numbers for cited pages within papers and other articles, although they should continue, as now, to provide the first and last page numbers of the whole article.

Bridging the quality gap

Sir — Between 1975 and May 2000, a total of 330 science and technology papers with at least one author from the University of Botswana (the only university in the country) have been published in journals with an SCI impact factor. Of these, 90% have been cited fewer than five times and 54% have never been cited (excluding self or co-author citations). The average number of citations per paper is 1.95. The current rate of publication averages one impact-factored paper per academic staff member every three to five years. These figures represent the bulk of the country's scientific output. Why is research productivity and impact so low?

Botswana is one of the more prosperous countries in the region, but the university offers little opportunity for good research. It is usually published in journals without impact factors and in conference proceedings. Until recently, few researchers seemed aware of the use of impact factors and citation analyses. However, with the equalizing effect of the Internet and ready access to abstracting journals and citation databases, we should become more active participants in quality research.

Developed countries can help others to raise the impact of their research and to bridge the quality gap. Instead of viewing postgraduates from developing countries as a short-term source of money, Western universities should see their training as a long-term contract to stimulate research and maintain collaboration once their overseas visitors are back in post.

Most PhD graduates in developed countries do postdoctoral stints to hone their research skills. PhD graduates from the Third World generally return to their own country as fully fledged lecturers with all the associated demands of teaching and administration, hence removing them from significant research activity.

Part of the problem is that senior administrators at universities in developing

countries seldom have scientific backgrounds and do not appreciate what it takes to do good science. A compounding problem is that — unlike the situation in developed countries, where scientists are trained and then recruited into academic institutions — in many developing countries staff are recruited as BSc graduates and then trained. Their early promise may not be realized. Universities in developed countries that train them seem willing to compromise standards, under pressure from their governments to increase income. Some Western academics apply double standards when asked, as part of an external peer-review process, to assess applicants for promotion in developing-world universities — recommending people they admit would not meet the criteria of the Western institution.

Every country has the capacity to contribute to the body of scientific knowledge. But for developing countries, without appropriate incentives and assistance from within and without, the quality gap will continue to widen.

Indra Riddoch

Faculty of Agriculture, University of Botswana, Private Bag 0027, Gaborone, Botswana

The public has (rightly) lost faith in scientists

Sir — The mishandling of BSE (*Nature* 408, 1 & 5; 2000) is not due to political or commercial interference in research. It is due to the failure of British scientists to live up to the traditional ethical standards of empirical science. In this, they are no worse than scientists in the United States.

Scientists everywhere have become dependent on governments and multinational corporations for funding. There still are scientists who live up to their responsibilities to serve the public, but they are now a powerless minority among those who call themselves 'scientists' because of their paper credentials.

If genuine scientists had not been an impotent minority, the British politicians and the commercial interests that pressured the politicians would have recognized the dangers of a public backlash from their actions.

As things turned out, the British public has so little faith in science that the current 'righteous indignation' of scientists fails to arouse very much public indignation. Concerned citizens no longer expect scientists to act to protect their health and safety. The public realizes it has to protect itself, with little help from scientists.

Irwin D. Bross

Biomedical Metatechnology Inc., 109 Maynard Drive, Amherst, New York 14226, USA

Speaking in too many tongues

Linguists are divided in their ideas about the origin of language.

Human Language and Our Reptilian Brain: The Subcortical Bases of Speech, Syntax, and Thought

by Philip Lieberman
Harvard University Press: 2000. 240 pp.
\$39.95, £26.50

Massimo Piattelli-Palmarini

Reminiscing about the beginnings of her career, the distinguished psychologist Eleanor Rosch once wrote: "I wanted my new field to be empirical, but not barbarically so." In his fifth book on the evolution of language, Philip Lieberman wants his field to be empirical, period. His central thesis is stated unequivocally on the second page: "Ultimately, human linguistic and cognitive ability can be traced back to the learned motor responses of mollusks." Such bluntness has to be put in the context of the success of Steven Pinker's 1994 book on language evolution, *The Language Instinct* (William Morrow), one of the best things to have happened in the dissemination of ideas in linguistics and cognitive science. Professor Lieberman probably felt obliged to write the ultimate anti-Pinker.

His thorn, from Pinker's rose, is that nearly half a century ago, at the Research Laboratory of Electronics at MIT, Noam Chomsky opened a new avenue into the study of language. Linguistic inquiry was to be part of the natural sciences, with a privileged link to biology. It would therefore welcome the inevitable consequence that linguistic hypotheses and explanations would be open to refutation or corroboration by data from a variety of domains — from traumatic speech impairments to congenital language deficits, from sign languages to the spontaneous creolization of pidgins, and much beyond. This new study of language was expected to open a privileged window on the workings of the human mind.

And many of us think it did. But an all-important proviso existed then, and persists today: the study of language is ultimately a biological science, but it must be conducted at a suitably abstract level of analysis. The symbolic representations and transformations that constitute a person's knowledge of their mother tongue must be investigated by amassing a huge variety of phonological, morphological, syntactic and semantic facts, from as many languages as possible. The brain scientists will have to know exactly what they are expected to find the neural bases of.

Lieberman totally neglects this proviso. To him, linguistics is applied evolutionary



"But wait a bit," the Oysters cried, "Before we have our chat."

biology. "In time, 'biological-linguists' working in an evolutionary framework will lead the way to new insights on the nature of language."

He claims to be in a position to offer a new kind of linguistics, but never attempts to tackle even one of the problems that have absorbed hundreds of linguists over the past several decades. He dismisses all their hard work at a single stroke, radically misquoting as his ally Ray Jackendoff, one of the most productive researchers in the field.

Jackendoff's idea, in the passage Lieberman refers to, was that any five-year-old child knows things about their mother tongue that certified linguists have been striving to understand for decades, so far with only limited success. This fact is explicitly presented by Jackendoff as the "paradox of language acquisition", which leads to the inevitable conclusion that "the human brain contains a genetically determined specialization for language" (from *Patterns in the Mind: Language and Human Nature*, Harvester/Wheatsheaf, 1993).

Admittedly, Jackendoff also candidly confesses to the limited success of his discipline. (What major scientific discipline will not admit that quite a lot is yet to be discovered?) Lieberman, using indirect discourse, drastically denatures this act of modesty: "Candid exponents of the algorithmic approach introduced by Chomsky such as Ray Jackendoff admit they cannot even describe the sentences of any human language." Such token of unfairness has a reason: those who believe language to be a practical ability, a skill based on a variety of

other skills, mobilizing huge, distributed functional complexes in the brain, have no truck with an "algorithmic approach".

In Lieberman's conception there is nothing specific about language. Syntax is merely the outcome of general patterns of motor control acting on the phonatory apparatus. The book deals at length with the transient linguistic impairments provoked by high-altitude hypoxia near the top of Mount Everest, or in hypobaric chambers. Who would deny that there are circumstances in which motor impairments and speech impairments occur together? The observation of drunkards would have been equally telling, although less exotic.

In a rather long quote, which appears twice, Lieberman endorses the idea that "thought is mental movement without motion". J. B. Watson's old behaviourist idea that syntactic comprehension consists of silent phonetic rehearsal (literally talking to oneself) is rediscovered, and occurs on so many pages that I could not even begin to list them.

I have just bemoaned Lieberman's liberal use of passages from authors in the other camp. The book also contains a few glaring factual mistakes that a good copy editor should have amended. But don't let me be misunderstood: although Lieberman is highly selective in his citations, he also correctly summons quotes and data from authors in his own camp. The idea that language is only a special chapter of general intelligence, coming out through sophisticated, but still generic, motor schemata, is very old and still popular. Lieberman has no shortage of real

book reviews

allies, in philosophy, anthropology, artificial intelligence, evolutionary psychology and the neurosciences. He can, and does, summon plenty of academically certified witnesses for the defence. None of them cares a jot about the problems, data and theories of Lieberman's so-called "algorithmic" linguistics. But indifference to, and ignorance of, linguistic research can too quickly turn into hasty indictments: "No evidence exists that ... the neural machinery regulating language differs fundamentally from that regulating other aspects of behavior. ... The search for the Universal Grammar is perhaps best regarded as a search for a holy grail."

But a central consideration must be taken into account. Lieberman has an endowed chair in a major university, and the monograph is part of a prestigious series from a prestigious publisher. Its very existence attests to a deeply unsatisfactory situation in large sections of this domain. Rosch's proviso has been too often neglected. The Liebermans of this world jettison *en bloc* the results of the real study of real languages, freely overgeneralizing from what they think they can observe in the brain, and in molluscs. And the other camp is not overly interested in reading about the alleged neuro-evolutionary bases of linguistic phenomena and mechanisms that they *know* cannot be right. With a few comforting exceptions, the long-sought, deep dialogue between linguistics and the neurosciences has yet to come. ■

Massimo Piattelli-Palmarini is in the Department of Linguistics, University of Arizona, Tucson, Arizona 85721-0025, USA.

Celebrating blurry boundaries

The Extended Organism: The Physiology of Animal-Built Structures

by J. Scott Turner

Harvard University Press: 2000. 235 pp.

\$47.50, £32.95

Steven Vogel

The snail's shell is an integral part of the snail, but the shell of the hermit crab is just clothing. There is no ambiguity here — the snail grows its shell, whereas the crab must shop repeatedly for empty shells that fit. Our hairs make up part of us, but the spider's web is just a structure that it builds. In this example, the distinction works less well. We may remain attached to our hairs, but the spider can (and does) reingest and recycle its fibroin, something we cannot do with our keratin.

Scott Turner begins his book by arguing that what we consider to be the outer boundaries of organisms are mainly habits derived



Community life: termite colonies operate communally built ventilation systems.

from perceptual accidents. He goes on to assert that such habits have led to an excessively restrictive view of what constitutes an organism, and that such a view has had unfortunate consequences for physiology. He bases his argument not on history, philosophy, cognition or our predilection for sharp dichotomizations, but rather on the weight of examples. With case upon case he shows how the sharp, traditional line between organism and external world often proves at least a nuisance and how, almost as often, we tacitly ignore it. And he concludes that our outlook on how organisms function would be empowered by drawing a more encompassing line.

If applied to a less eclectic field, or if it were the account of a less eclectic scientist, Turner's case-by-case approach might be tedious. But few readers of this book will fail to be fascinated by the examples. Turner's tales of the subtle ways organisms capitalize on the opportunities afforded them by their physical and chemical surroundings provide more than ample reason to read the book. Although no area of biology brings into its purview more diverse elements of the physical sciences than physiology, few physiological accounts consider both the present range of non-biological factors and the full range of plants, animals and microorganisms in the way that Turner does.

On the physical (or mathematical) side, one encounters diffusion, fractals, redox potentials, acoustics, thermodynamics and chemical kinetics, hydrostatics and dynamics, convective and radiative heat transfer, soil mechanics, surface tension, climatology and control systems — each with clear explanations adequate for their concomitant biological stories.

The biological side is no less diverse. One learns about bioconvection and the way

biological and physical factors interact to produce large-scale order in populations of swimming microorganisms; about plasmons and the other external bubble-lungs used by aquatic but air-breathing insects and spiders; about the way gall midges manipulate the surface temperatures of solar-heated leaves; and about the thermoregulatory tricks of the aerial bee colonies and subterranean termite colonies that operate communally built ventilation systems.

Turner gives particular attention to two general topics, each providing a set of elaborate cases where physiology must take account of "external" structures. Burrowing appears in multidimensional splendour, from its palaeontology to such things as the mechanics of digging, electron acceptors at different depths and in different substrata, the necessary sensitivity of earthworms to soil-water potential, and the use of burrows as feeding devices. The latter focuses on how lugworms — as cleverly adapted as they are consummately ugly — cultivate microorganisms in their burrows, and provides a splendid mix of zoology, microbiology, physical chemistry and fluid mechanics.

The other general topic is animal acoustics and the way external structures help small animals produce loud sounds whose characteristics are attuned to their functional roles, such as the cricket that improves the emission of its song by cutting a hole in a leaf that then serves as a baffle. The acoustic account then returns to burrows, specifically those of mole crickets and the way in which the shape of these residences allows them to work like different musical instruments — and to the feedback schemes by which mole crickets, like musicians, listen and retune.

Turner's views are less a radical revision than a reminder or rejoinder. Indeed, the argument might be pressed still farther. One

can equally well take a narrower view of the organism-surroundings boundary, limiting the organism to intracellular material that participates in the continuous processes of breakdown, excretion, resynthesis and resupply. Hair, horn, exocuticle, even wood then join the domain of burrow, nest and cocoon. The zone of context-dependent definition of an organism extends more certainly inwards than further outwards towards 'Gaia' and superorganisms, with their metaphorical and metaphysical baggage; although, to be fair, these latter get only brief mention in the book.

The attempt to wrap a larger picture around the rich complexity of organisms takes a bit of shoe-horning and discrete tucking in of loose shirt-tails — but one could make the same comment about Darwin's *Origin*. As Mark Twain put it in *Huckleberry Finn*: "There was things which he stretched, but mainly he told the truth."

Steven Vogel is in the Department of Biology, Box 90338, Duke University, Durham, North Carolina 27708-0338, USA.

Observations of a travelling naturalist

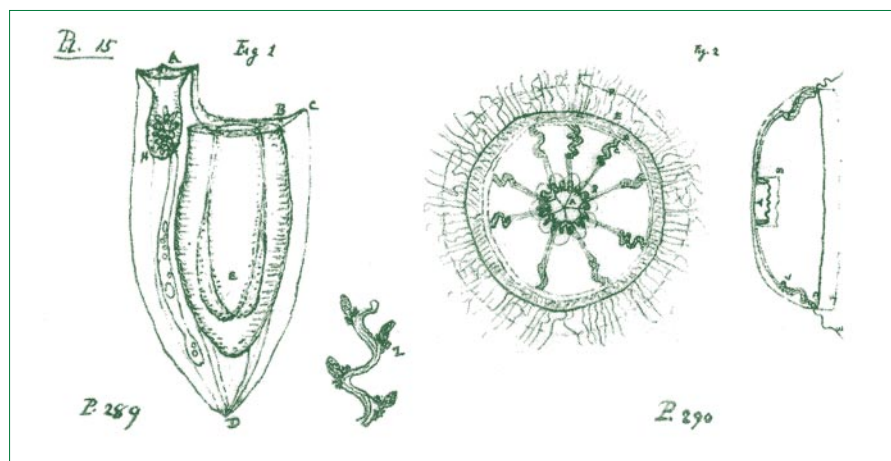
Charles Darwin's Zoology Notes & Specimen Lists from H.M.S. Beagle

edited by Richard Darwin Keynes
Cambridge University Press: 2000. 428 pp.
£95, \$150

Janet Browne

Even though Charles Darwin's experiences during the *Beagle* voyage have been endlessly — and fascinatingly — explored, there is much still to discover. Whereas current biographies try to do justice to the full adventure, and specialist volumes pick through Darwin's key moments, it can come as something of a shock to realize that there are highly significant manuscripts still languishing virtually unread in the archives of Cambridge University Library.

Prime among these are Darwin's zoological notes, the record that he compiled day by day as he travelled around the world. The publication of his complete correspondence from that period provides stunning insight into the larger world of early-nineteenth-century travel and natural history. Darwin's notes described his observations and dissections of marine organisms, invertebrates, birds, reptiles and mammals. These notes were additional to his journal, and to the pocket-sized field notebooks that he carried with him on expeditions ashore, and the lists of specimens he made. The zoological notes contain the information about nature in the raw that he wanted to capture for



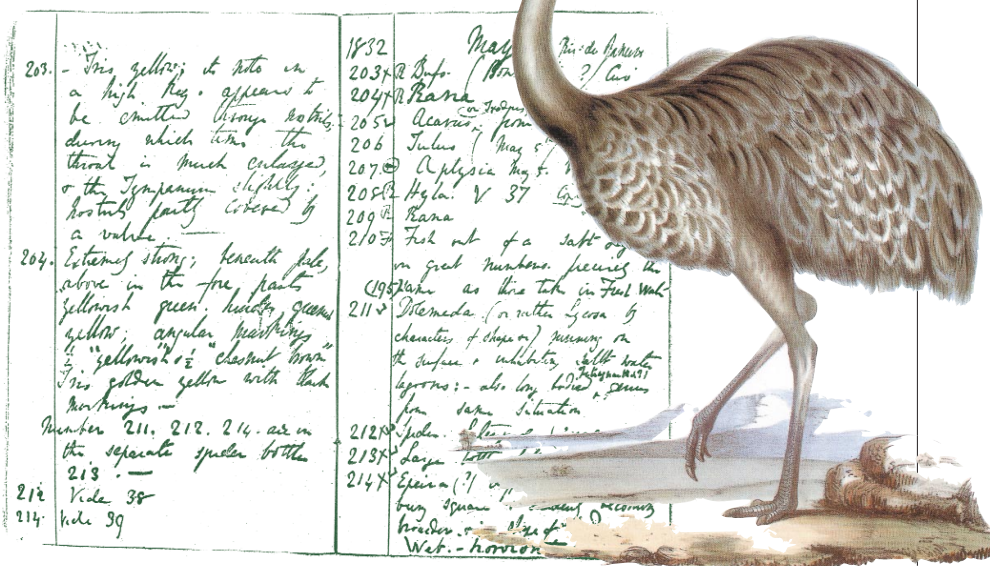
posterity — the habitat descriptions, records of behaviour, specimens' stomach contents, animals' appearances in life, and observations under the microscope that would be essential when he arrived back in England.

These provided the real backbone to Darwin's understanding of the biological kingdom, and show that even on the voyage he was grappling with momentous questions about the definition of life, individuality, reproduction and taxonomic relationships. Darwin's zoological notes made during the *Beagle* voyage, in fact, reveal him as a true biologist, working and thinking at levels of intensity that were afterwards somewhat masked by the informal, self-deprecating style of his *Journal of a Naturalist*. But they would emerge in their most powerful form in the *Origin of Species* in 1859.

Richard Darwin Keynes provides biologists with a real treat by making these important manuscripts accessible for the first time. Darwin's writing style was vivid and accurate. Here and there he added the kind of impromptu sketch that a field naturalist would recognize immediately as the best means to jog the memory after a specimen

has disappeared into its jar or travelling cask — the leading characteristics of a puzzling water flea, perhaps, or the internal anatomy of coralline algae, the organisms that first caught Darwin's imagination as a medical student working with his mentor Robert Grant on the banks of the Firth of Forth in Scotland.

One ghost that will be laid to rest is the extent to which Darwin might be thought to have been an amateur naturalist on this voyage. It is perfectly clear that he was a superbly skilled observer who intended making the most of the unrivalled opportunities presented to him. In his notes he carefully compared new organisms with descriptions of comparable species in the authoritative textbooks available to him — the comprehensive library on board the *Beagle* was a great asset to his researches. He was meticulous in dissecting under the microscope, performing small experiments and recording the necessary details. He recorded his questions and possible fields for future research when he returned to England. And it should be said that for many years after the *Beagle*'s return Darwin planned to classify his invertebrate collection himself.



Keynes has identified most, if not all, of the animals Darwin discusses, and these helpful identifications add greatly to the volume's value. Providing these identifications must have been an enormous undertaking.

The text is arranged in semi-facsimile, so that the sequence of Darwin's researches, and much of his excitement, is retained. The steps in his self-education in the biological sciences become much clearer. Nor did Darwin confine himself to strictly zoological observations. The phosphorescence of the sea, for example, transfixed him with its beauty, although he made sure also to strain a pint pot of the luminous water through fine gauze and put it under the microscope. "In the water were some minute Crustaceae of the genus *Cyclops*. I should not be surprised if these added to the effect." Many of his descriptions of land and maritime scenery are very fine. After the *Beagle* returned, Darwin used a number of these passages to enhance his *Journal of a Naturalist*.

The volume includes a scholarly and sympathetic introduction, and transcripts of the relevant catalogues of species that Darwin made during the expedition. Keynes has previously rendered signal service to Darwinists of all persuasions with his account of the *Beagle Record* and by transcribing anew Darwin's *Beagle Diary*. Here is another distinguished contribution that will illuminate this special region of Darwin's heart. ■

Janet Browne is at the Wellcome Trust Centre for the History of Medicine at UCL, University College London, 24 Eversholt Street, London NW1 2AD, UK.

Science in culture

Armed with anatomy

A nineteenth-century anatomy lesson for artists.

Martin Kemp

Occasionally, a single image sums up a long-standing enterprise in a way that no other does. François Sallé's painting *Anatomy Class at the École des Beaux Arts* perfectly encapsulates the professional instruction in anatomy that had been enshrined in the statutes of art academies since the sixteenth century.

The painting (below) shows Professor Mathias-Marie Duval, author of *Cours d'anatomie* (1873), demonstrating key features of superficial anatomy on a well-muscled man in working-class trousers, who is standing with confident poise on a podium. This particular lesson is concentrating on the arm. On the table are the bones of the arm and shoulder, while an outline drawing on the backboard shows a scapula and humerus from the rear. Duval extends the model's arm for scrutiny, having asked him to clench his fist to increase muscle and tendon definition. And he may be demonstrating the ingenious mechanism behind the rotation of the wrist, in which the radius and ulna are twisted across each other, and which involves the action of the biceps.

Most of the attentive audience of artists are seated within the horseshoe terrace of benches — arranged like one half of the traditionally round or elliptical anatomy theatres common in medical schools. As was the norm, there are no women present in the irredeemably masculine confines of the life room; female anatomy was only studied for those features that men did not share.

In the midst of the audience stands a vigorous écorché statue, probably a plaster cast from the body of a criminal which had been

flayed to disclose the muscles below the skin and subcutaneous fat. Carefully posed — one arm raised and one lowered, one fist clenched and one hand open, one leg stretched and the other bent — the écorché would have served as an anatomical exemplar outside the professorial lecture and during the very occasional dissection. It is likely that the upright cabinet beside Duval houses a full human skeleton; and the mounted skeleton of the bird among the jars and flasks on top alludes to the professor's interest in comparative anatomy in the tradition of the French zoologist Georges Cuvier.

Such anatomical teaching reflected the academic conviction that the human body could not be portrayed effectively as an expressive vehicle if the artist did not know how the mechanisms of motion and expression worked — from the inside out. The first Renaissance art treatise, Leon Battista Alberti's *On Painting* in 1435, set an anatomical agenda that was to endure for more than 400 years:

"First ... sketch in the bones, for, as they bend very little indeed, they always occupy a determined position. Then add the sinews and muscles, and finally clothe the bones and the muscles with the flesh and skin ... There will perhaps be some who will raise an objection ... namely that the painter is not concerned with things that are not visible. They would be right to do so except that, just as for a clothed figure we first have to draw a naked body beneath and then cover it with clothes, so in painting a nude, the bones and muscles must be arranged first, and then covered with appropriate flesh in such a way that it is not difficult to perceive the positions of the muscles."

Exhibited at the annual Salon of the Academy in Paris in 1888, where it was awarded a gold medal of the third class, Sallé's large canvas exults in the tradition even as it stands at its point of decline. Anatomy, like the discipline of perspective, was losing its grip on artists' imaginations. The creative vanguard of painting was moving decisively away from the canons of naturalism that had dominated for half a millennium.

The masterpiece by Sallé (c.1839–99), an artist little known today, was purchased directly from the Salon for the Art Gallery of New South Wales and immediately shipped to Australia. It remains less familiar than it deserves to be. ■

Martin Kemp is in the Department of the History of Art, University of Oxford, 59 George Street, Oxford OX1 2BE, UK.

Anatomy Class at the École des Beaux Arts is on view at the Hayward Gallery, London, until 14 January 2001, in the exhibition *Spectacular Bodies* (see *Nature* 408, 140; 2000).



CHRISTOPHER SNEE/ART GALLERY OF NEW SOUTH WALES

The all-chemist

Linus Pauling set the agenda for a century of chemical research.

Gautam R. Desiraju

All of chemistry is a fugue between the statics of structure and the dynamics of reaction, and at its very core lies the covalent bond. Physics and chemistry went hand in hand until the early decades of this century, when the ionic bond followed from Niels Bohr's model of the atom. But it was only when the idea of electron sharing was accepted and quantified that the chemical bond came into its very own. Chemistry had no further need of physics for the rest of the century. This extrapolation from physics to chemistry and the articulation of chemistry as an independent subject was the handiwork of a single individual. Linus Pauling ranks with Galileo, Da Vinci, Shakespeare, Newton, Bach, Faraday, Freud and Einstein as one of the great thinkers and visionaries of the millennium. Truly he was not of this age, but for all time.

Pauling's ingenuity and awesome intuition permeated quantum mechanics, crystallography, biology, medicine and, above all, structural chemistry. He was curiously silent on organic chemistry, even after enunciating its most fundamental feature, namely that a sharing out of electrons evenly among equivalent energy states — the hybridization of bond orbitals — explains the tetrahedral valences of the saturated carbon atom. His influence on inorganic chemistry was profound. Hybrid orbitals were spectacularly successful in explaining the magnetic properties of transition-metal coordination compounds and the structural vagaries of organometallic complexes.

The late nineteenth and early twentieth centuries saw the growth of the molecular paradigm in the hands of the grandmasters. August Kekulé, Adolf von Baeyer and Richard Willstätter, among others, developed the idea of the simple molecule, and the inorganic chemist Alfred Werner took this forward to the extended molecule in his description of coordination compounds. Pauling's early work brought all of this within a common model and in so doing effected a grand unification of molecular structure.

In retrospect, it appears that chemistry was waiting for Pauling. His chapters in *The Nature of the Chemical Bond and the Structure of Molecules and Crystals*, one of the most influential books of the twentieth century, proceed almost ruthlessly through interatomic distances, electronegativity, ionic, covalent and van der Waals radii, aromaticity and the structure of benzene, multiple bonds, electron-deficient substances, the



All-conquering: the next generation of chemists must escape from Pauling's shadow.

metallic bond and the hydrogen bond. They were a route map for chemical research for the rest of the century.

Pauling did not work directly on the dynamics of chemical reactions, but the implications of his work on the subject are all too apparent, for unless one understands the molecule at rest, one cannot begin to comprehend it as it stirs. Chemistry grew and prospered simply by proving time and again that Pauling was correct in just about all his conjectures, for he projected with unerring accuracy into the future with only about 0.01% of today's structural information.

Chemistry, then, is utterly different from physics and biology in its dependence, at a primal level, on just one scientist. This raises some awkward questions. Did Pauling's influence modify, divert or even stunt the development of other, perhaps still unexplored or unidentified, branches of the subject? Are chemists introverted and cautious because much of their research flows from a single stream of consciousness, making big imaginative leaps unnecessary? Does chemistry lack the high drama of physics and the glamour of biology because these were appropriated by one larger than life individual?

These questions are not completely frivolous. Pauling elevated the molecule to the high altar and it became the delimiter of all the important properties of a substance, to the extent that there was no world outside it. This could account for the relatively late take-off of supramolecular chemistry, nearly

70 years after Emil Fischer enunciated his lock-and-key principle and Paul Pfeiffer described the first molecular complexes. Again, the central position given to the covalent bond could account for solid-state chemistry's struggle to join the mainstream. Even today, computational chemistry is largely concerned with discrete molecular systems. As for structural chemistry, the phenomenon of hydrogen bonding did not allow for weak donors and acceptors for decades, mostly because Pauling had said that only strong donors and acceptors could form these bonds.

To paraphrase Jean-Marie Lehn, chemistry is all about diversity, whereas biology revels in complexity. Pauling's work does not fade with ever-increasing diversity, for it is too fundamental with respect to structure. It is as chemistry moves from the molecule towards the study of complex systems that chemists will have to think beyond Pauling. Very short timescales, very large length and dimension scales and the intersection of chemistry with materials science and biology will hold chemists' attention. Chemistry will become a different subject as it transforms from a unitary to a diversified science. Will it ultimately delocalize and dissolve into other disciplines? Only time will tell, but if it does, it would only be because its very identity depended for so long on the contributions of a single individual. ■

Gautam R. Desiraju is in the School of Chemistry, University of Hyderabad, Hyderabad 500 046, India.

Brain drain

Why civilizations fall silent.

Frederik Pohl

As Docent Wilfram's actual name, which summarized his complete medical and professional history, was very long, his friends simply called him "Wilf". There weren't very many of those friends left, though, because the number who were still in any real sense alive — that is, not frozen, cremated or in machine storage — dwindled year after year.

That was natural enough. Wilf had been birthed in 2734 and was therefore now a hundred and seventy-four years old, and, although genetics, microrobotic surgery and easily available custom-grown transplants had given most human beings a life expectancy undreamed of in earlier times, they couldn't keep a person from feeling old — even when Wilf's housemind, Jerel, was giving him news which, at one time, Wilf would have considered exciting. "Message?" Wilf repeated. "An ET message?"

"Definitely ET," his housemind assured him, "though its content, if any, is unknown. It was received by the thousand-hectare radio telescope in Trojan-Uranus."

"Yes," said Wilf, yawning, "Well, you never know: send it along for analysis. I think I'm going to go to sleep."

As he headed for his bedroom he had already nearly forgotten the news. Messages from space had not been headline news for a long time. The first message — that is, at least, the first radio signal which could not possibly be natural in origin and thus had to be the product of extra-terrestrial intelligence — had been detected as early as 2063, and as detection facilities improved over the years 37 others had been logged. They came from all over the sky, some a few score light-years away, others more than a thousand.

But, in spite of the best efforts of humanity's increasingly capable computers, no message had ever been decoded, and there was increasing doubt that there was anything to decode. Nothing but the inevitable radio leakage from any high-tech civilization.

That was disappointing, but there was worse. Although 38 extraterrestrial radio sources had been detected, only 11 were still on the air. The rest had gone silent and stayed that way. Why?

That was the part Wilf didn't like to think about. It seemed that most high-tech civilizations lasted only some centuries. Then something happened to them. What that something might be no one could say, but



there was one unwelcome theory that would not go away. Any civilization that reached the point of large-scale radio emissions was likely at the same time to be developing weapons of mass destruction. And that, it seemed, was a death sentence.

As Wilf limped toward his bedroom, the housemind spoke again. "Docent Wilf? I ask again, may I fix that limp for you?"

"It isn't worth the trouble!" Wilf said.

"It is no trouble," the housemind persisted. "Although, to be sure, it would be more efficient for you to have yourself machine-stored now, and then such problems would not arise."

"Yes, I know," Wilf said testily, throwing himself on his bed. Generally it made more sense to follow a housemind's advice, as their machine minds were far more capable than any human's — especially that of a human who still clung to his organic body. He was also quite sure that Jerel knew something it wasn't telling its master. It might be, he thought, about his own life expectancy. If

you were going to have yourself copied into a computer program it was a good idea to get it done while you were still alive.

After actual death, even a short time after death, there was a certain degradation of the data. And there really was no reason to put it off. You lost nothing in machine storage. Indeed, you gained a world — any kind of world you wanted! You could create any virtual reality you liked and live in it as long as you chose, and when you tired of that you could create a different and even better one.

Nearly all Wilf's age-cohort had long since taken the step themselves, and when they talked to him about it — when they bothered at all to talk to any person who was still flesh and blood — they unanimously described it as the closest thing any non-believer could get to a heaven of his own.

Wilf sat up suddenly, opening his eyes. "Jerel!" he called. "Show yourself! I want to talk to you."

"Yes, Docent Wilf?" The housemind obediently presented itself as the hologram of an ancient English butler, standing attentively a couple of metres away.

"You've thought this through, haven't you? Those other civilizations didn't wipe themselves out in wars, did they?"

The housemind's expression clouded. "Why do you ask, Docent Wilf?"

"I ask you because you've got a better mind than I have."

"In certain areas, perhaps," the housemind agreed.

"They don't all die, do they? They just put themselves in machine storage. And then they've got nothing to worry about, ever — not hunger, not illness, certainly not death. Not even unrequited love, because if their love object isn't in a requiring mood they simply simulate her when she is. And they have such a grand time they don't bother with anything else."

"Generally speaking, no," the housemind agreed.

Wilf laughed. "Of course not," he said. "And neither will we, will we? There won't be any more signals to leak to the rest of the Galaxy! And so as soon as the rest of us are in machine storage, Earth will fall silent, too." ■

Frederik Pohl has been active as a writer, editor and agent since the 1930s, and in 1992 the Science Fiction and Fantasy Writers of America made him a "Grand Master" of SF in recognition of his contribution to the field. His next book is Chasing Science (Tor), a rhapsody on the pleasures of science considered as a spectator sport.

Gaining light from silicon

Leigh Canham

Silicon lasers would help computers operate faster by replacing electrical connections with optical ones. The trouble is, normal silicon doesn't glow. Densely packed silicon nanocrystals could be the answer.

Silicon technology is so powerful that it pervades our everyday lives. Silicon chips are in our homes, our cars and even in some people's bodies. These micro-circuits help us wake up on time, surf the Internet, and keep elderly hearts beating in a rhythmic fashion. Indeed, the scientific literature contains more than 250,000 papers relating to silicon. But it has been known since the 1960s (shortly after the invention of the integrated circuit) that bulk silicon is extremely inefficient at emitting light, and so could play only a minor role in optoelectronics — the high-speed future of electronic circuits. So, to make lasers and high-speed telecommunications devices, scientists turned to more complex semiconductors, such as gallium arsenide and indium phosphide. These are good at emitting light but are more expensive than silicon and are hard to integrate into silicon microchips. If an all-silicon laser could be created it would revolutionize the design of supercomputers and lead to new types of optoelectronic devices.

The findings reported by Pavesi *et al.*¹ on page 440 of this issue suggest that silicon lasers may yet emerge. If this paper had appeared more than ten years ago, it would have met with disbelief in the semiconductor community, but today it should cause more excitement than scepticism. This is because our perception of silicon has changed over the past few years. It is now known that the properties of silicon are sensitive to its structure at the nanometre scale.

In particular, silicon nanocrystals and highly porous silicon can emit red, green and even weak blue light when stimulated by light of shorter wavelength (Fig. 1). The efficiency of silicon nanocrystals (measured in photons emitted per incident photon) can exceed 1%, which is 10,000 times better than bulk silicon. The origin of visible light from porous

silicon has been hotly debated, but for light-emitting diodes (LEDs) made from porous silicon, the efficiency (the ratio of photons emitted per injected electron) has also risen steadily over the past decade², and now stands at about 1%. This is still too low for practical devices, but is closer to the efficiency of a few per cent needed for integrated light-emitting displays.

There is still some way to go before LEDs made from porous or nanocrystalline silicon are commercially attractive. One of the problems is the nature of the optical bandgap — the gap in the electronic energy levels that controls light emission. Although silicon nanocrystals have an energy bandgap wide enough to produce visible light (2–3 electron volts rather than 1 electron volt in bulk silicon), they retain some of the 'indirect bandgap' nature of normal silicon. This means that light emission can occur only when accompanied by an additional process that transfers energy. As a result, light production in these nanocrystals is slower (microseconds) than that of a 'direct bandgap' material such as gallium arsenide (nanoseconds). This means that processes that generate heat rather than light can compete with the emission of photons and severely limit light output. It also means that silicon LEDs can be switched on and off only at rates of 1 megahertz, rather than the 100 megahertz or 1 gigahertz rates needed for high-speed optical connections³.

Theoretical calculations show that the size of silicon nanocrystals will need to decrease to 1 nanometre to emit blue light as efficiently as compound semiconductors such as gallium nitride. So many researchers are investigating ways of making even smaller and more uniform silicon nanocrystals⁴. Others have tried sandwiching layers of luminescent nanocrystals between tiny mirrors

to make an optical cavity that can amplify the light emission. For porous silicon, this has led to significant narrowing of the emission spectrum, which is one of the features of 'coherent' laser-like light (when all the photons are in step with one another). But so far there has been no improvement in the speed of emission, which keeps the intensity of emitted light low. The best solution would be to convert the normal spontaneous luminescence process within silicon nanocrystals into the sort of 'stimulated emission' associated with lasers. In a laser, each emitted photon stimulates the emission of another photon with the same frequency, resulting in light output being both amplified and coherent.

Which brings us to the remarkable findings of Pavesi and co-workers¹. They have created a densely packed assembly of silicon nanocrystals, 3 nanometres wide, embedded in an oxide matrix. This type of silicon nanostructure, like porous silicon, was first produced some time ago⁵, but its luminescent properties lay dormant until recently. In the experiment of Pavesi and colleagues, the nanocrystals form a thin layer, just below the surface of an oxidized silicon wafer. The authors use two standard methods for measuring the 'optical gain' (when light output exceeds light input), and obtain a remarkably high gain in both cases — comparable to that from assemblies of compound semiconductor nanocrystals, such as gallium arsenide.

The authors attribute this high gain to the extremely high density of silicon nanocrystals in the structures, because their estimates of the gain per nanocrystal are much lower than those of direct bandgap semiconductors. Demonstrating optical gain is a crucial step towards making a silicon laser, but it is not the end of the story. A true laser has to produce coherent as well as amplified light,



Figure 1 First light for silicon. Ten years ago, researchers managed to get nanocrystalline silicon to emit red light by stimulating it with an argon-ion laser beam (green light). Scientists hope that by the year 2010

they will be able to use a nanocrystalline silicon laser instead of the argon one. Such tiny silicon lasers would revolutionize the growing field of optoelectronics.

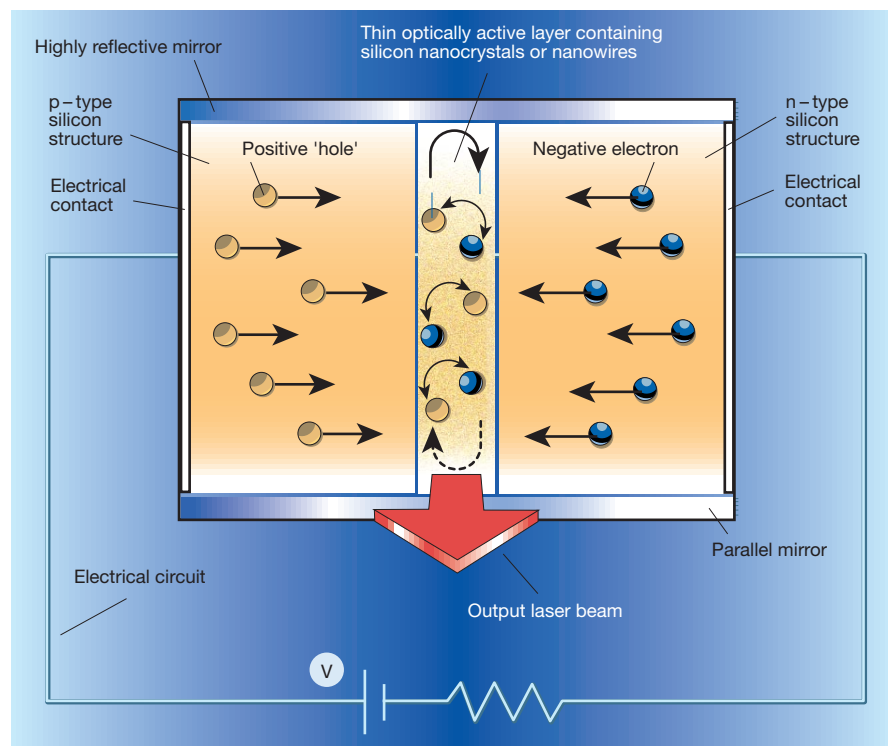


Figure 2 How an electrically driven silicon laser might work in the future. Electrically driven emission of light from semiconductors is the basis of tiny light-emitting diodes (LEDs) and lasers. Traditional semiconductor LEDs are formed from p-type and n-type semiconductors, which donate positively charged 'holes' and negatively charged electrons, respectively, when a voltage is applied across the structure. Recombination of an electron and hole, within the semiconductor, produces a photon and leads to the emission of light. If the efficiency of light emission is high enough and the whole structure is placed between two highly reflective mirrors, the LED can be turned into a miniature laser. Silicon — the semiconductor of choice for the electronics industry — used to be considered a poor emitter of light, but Pavese *et al.*¹ have demonstrated good optical emission from a layer of silicon nanocrystals stimulated by pulses of ultraviolet light (not electrons). Now the challenge is to electrically stimulate these nanocrystals into producing a beam of laser light.

to achieve the narrowness and intensity required of a laser beam.

Why has optical gain not been seen in the much more intensely studied porous silicon? One difference is clear from the optical absorption spectrum of the nanocrystals, (see Fig. 1 on page 440), where there is a feature that has not been reported for porous silicon. Pavese *et al.* suggest that their gain is a consequence of the high quality of their nanocrystal–oxide interface, which has many 'surface states' that emit light per nanocrystal. By contrast, in unoxidized porous silicon the luminescence process does not involve such surface states, and is much more dependent on the size and shape of the nanocrystal. As with lasers made from compound semiconductor nanocrystals, high optical gain would require a much more uniform nanocrystal size distribution than etching or Pavese *et al.*'s technique has achieved.

What next? To build a silicon laser these silicon nanostructures must be made to work inside an optical cavity, in which mirrors can bounce the light back and forth until it becomes coherent (Fig. 2). The other key feature missing from Pavese and colleagues' system is the electrical stimulation of light

emission. Replacing the existing optical stimulation with an electrical process is essential for silicon lasers to be integrated easily into electronic circuits (Fig. 2). Here,

Gene regulation

Local or global?

Shelley L. Berger

Modifications to histone proteins were thought to act only locally to control the expression of single genes. But the finding of such changes across the whole genome brings that view into question.

The US congressman Thomas 'Tip' O'Neill once famously said: "All politics is local". In the same vein, environmental groups have adopted the slogan "Think globally, act locally". In biology, a similar sentiment is thought to be relevant to gene expression. Here, it is held that changes in the structure of chromatin — a compact, compressed form of DNA — occur only 'locally', allowing the regulation of individual genes. But it seems that this picture is too simplistic, for, on page 495 of this issue¹, Vogelauer

and colleagues describe their discovery of genome-wide, 'global' chromatin modifications. The functions of these modifications remain something of a mystery.

The basic structural unit of chromatin is the nucleosome, which consists of a stretch of DNA wrapped tightly around a core of histone proteins. This compressed form makes it hard for transcription factors, which regulate gene expression, to gain access to target genes in the DNA. However, enzymes that add or remove phosphate, acetyl or methyl

attention may focus on the interconnected nanowire structure typical of porous silicon, rather than silicon nanocrystals, because the electrical properties of the silicon nanowires have been crucial to achieving high LED efficiency². Future studies of optical gain should also be carried out below room temperature, because the temperature dependence of these processes can yield considerable insight into the gain mechanism, and lasing is normally first achieved at low temperatures, where competing non-luminescent processes are less effective.

The work of Pavese *et al.*¹ and others over the past decade has only recently changed our view of silicon as a poor emitter of light. But if some astrophysicists are correct then nanocrystalline silicon has been emitting light throughout the galaxy for billions of years^{6,7}. They argue that the properties and cosmic abundance of silicon could account for the emission of red light seen by astronomers in clouds of interstellar dust. So silicon nanocrystals may be rather more abundant than previously thought, though it will be some time before such natural structures can be closely examined. In the meantime, once the results of Pavese *et al.* have been reproduced, they will, I'm sure, be seen as a milestone in attempts to develop silicon-based optoelectronics. It appears that we can still teach the 'old dog' of semiconductors a few tricks. It just needs to be restructured on the nanoscale.

Leigh Canham is at the Defence Evaluation Research Agency, St Andrews Road, Malvern WR14 3PS, UK.

1. Pavese, L., Dal Negro, L., Mazzoleni, C., Franzo, G. & Priolo, F. *Nature* **408**, 440–444 (2000).
2. Gelloz, B. & Koshida, N. *J. Appl. Phys.* **88**, 4319–4324 (2000).
3. Goodman, J. W. *et al. Proc. IEEE* **72**, 850–866 (1984).
4. Holmes, J. D. *et al. Science* **287**, 1471–1473 (2000).
5. Nesbit, L. A. *Appl. Phys. Lett.* **46**, 38–40 (1985).
6. Ledoux, G. *et al. Astron. Astrophys.* **333**, L39–L42 (1998).
7. Zubko, V. G. *et al. Astrophys. J.* **511**, L57–L60 (1999).

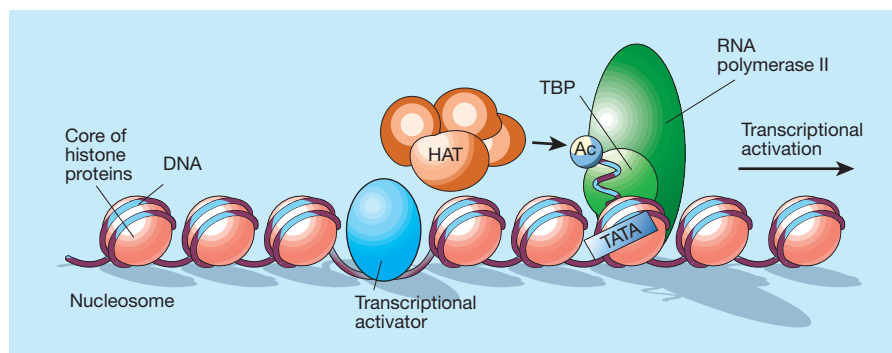


Figure 1 A commonly held view of gene regulation. Local modification of histone proteins, around which DNA is wrapped, is thought to allow the regulation of gene expression on a gene-by-gene basis. A histone acetyltransferase (HAT) adds acetyl groups (Ac) to histone proteins near the TATA box — a DNA sequence near the point in a gene from which transcription starts. This somehow allows the target gene to be transcribed into messenger RNA. The HAT is part of a protein complex that is recruited by physical interactions with a DNA-bound transcriptional activator. RNA polymerase II and the TATA-binding protein (TBP) are two of the key protein complexes involved in transcription.

groups from histone 'tails'^{2,3} somehow affect chromatin structure in such a way that genes can be turned on or off. The best characterized of these enzymes, which in effect act as transcription factors themselves, are the histone acetyltransferases (HATs), which add acetyl groups to the histone tails, and their opposite numbers, the histone deacetylases (HDACs). In a commonly held view, these enzymes alter gene expression by acting locally, gene by gene, to modify histone proteins near the regulatory 'promoter' regions of the genes.

Three observations led to this model. The first nuclear HATs⁴ and HDACs⁵ to be identified were known previously as promoter-specific transcription factors. Then it was discovered that both HATs^{6,7} and HDACs^{8–13} are components of multiple-subunit complexes that are recruited to specific promoters by interaction with DNA-bound transcriptional activators and repressors. And early findings^{14–16} suggested that both HATs and HDACs are strictly limited in their method of action. They seem to work on histone proteins in only one or two nucleosomes, found next to a 'TATA box' — the site in a promoter region to which the main gene-transcribing enzyme, RNA polymerase II, binds^{14–16}. A pleasing model emerged in which chromatin-modifying factors, such as HATs and HDACs, are recruited to specific promoters to regulate transcription directly (Fig. 1).

But Vogelauer *et al.*¹ have discovered that this is not the whole story. They have found that the best characterized of the histone-modifying enzymes have a more global role, apparently allowing a rapid cycle of acetylation and deacetylation throughout the genome. The authors use a powerful technique, chromatin immunoprecipitation, that allows *in vivo* examination of proteins and their modifications at specific points in chromatin. They find that acetylation occurs ubiquitously throughout a large chunk of the genome of the baker's yeast *Saccharomyces*

cerevisiae. The level of global acetylation is slightly higher than that seen at the ends of chromosomes (telomeres), which are transcriptionally silent. Moreover, at every point examined in the yeast genome, the amount of acetylation could be lowered by deleting a particular HAT, or raised by deleting a particular HDAC. The implication is that the state of acetylation of a genome is in constant flux. It is higher at promoter regions, but this seems to be superimposed over the global pattern.

A possibly related phenomenon was recently reported by Cohen *et al.*¹⁷, who used whole-genome microarray analysis in yeast to examine the relationship between the expression of 'linked' genes. They uncovered interesting correlations in the patterns of expression of adjacent genes that are normally regulated in different ways. This suggests that patterns of expression of large domains of genes underlie gene-specific patterns.

Two possible functions for the 'base' level

of acetylation discovered by Vogelauer *et al.*¹ come to mind. The first is that it helps to turn transcription on or off, priming or dampening transcription (Fig. 2). This would mean that chromatin, like an idling truck, is always ready to go. Just a bit more acetylation at a particular gene's promoter is needed to turn that gene on. Then only a little deacetylation is needed to return it quickly to the idling state. A role in priming would help to answer the thorny question of how transcriptional activators penetrate the chromatin barrier to reach target genes. However, Vogelauer *et al.* propose, and provide some evidence in support of, a dampening model only — that the global acetylation pattern allows chromatin to return quickly to the idling state after activation. But of course both may be true, and could be working together.

A second hypothesis is that these global modifications have a totally different role, perhaps not related to transcription at all. For example, when DNA is replicated as part of the cell-division cycle, it may be necessary to 'loosen' all the chromatin to allow the replication machinery access to the entire genome (Fig. 2). In this case, one would predict that the global acetylation pattern changes during the cell cycle. Perhaps higher levels of global acetylation are seen during the DNA-replication phase, with lower levels occurring during cell division itself, when DNA is more condensed. Another role for global acetylation might be to allow the DNA-repair machinery access to the genome, to 'shop' for mutations that need fixing (Fig. 2). Other functions can be envisaged.

An unanswered question in this field is how changes to histone proteins translate into changes in the level of gene expression. Two possibilities are in favour. First, changes to histone proteins may lead directly to physical changes in nucleosome or chromatin structure, making it easier for other tran-

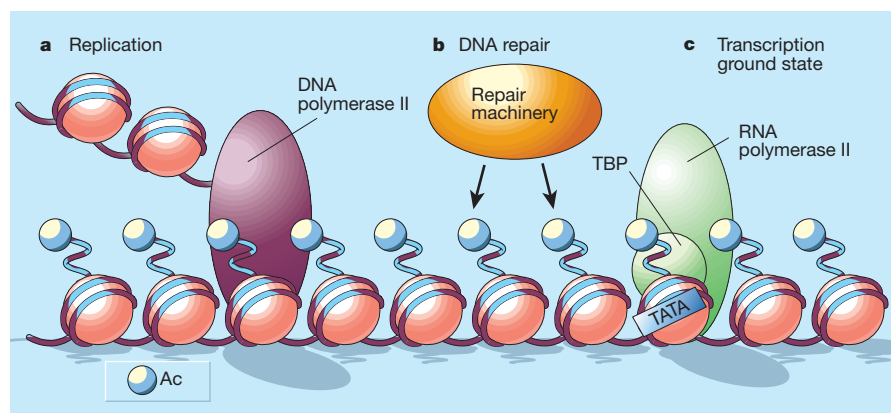


Figure 2 Processes that might be regulated by the base level of histone acetylation discovered by Vogelauer *et al.*¹ throughout the *Saccharomyces cerevisiae* genome. Such global acetylation, brought about by the histone acetyltransferases and histone deacetylases (not shown), might somehow loosen the entire genome, allowing access of: a, the DNA-replicating enzyme DNA polymerase II; or b, DNA-repair enzymes. c, Vogelauer *et al.*¹ offer evidence in support of another model. They propose that the base level of acetylation amounts to a transcriptional 'ground state' (the pale green shading denotes low RNA polymerase II activity), allowing target genes to be turned on and off rapidly. Ac, acetylation.

scription factors to gain access to target genes. Alternatively, the modifications may create a special pattern that provides a surface to which relevant factors can bind. Investigation of the global modifications discovered by Vogelauer *et al.*¹ may offer further insight into this problem.

During the past 40 or so years of molecular biology, researchers have generally used single genes to study genomic regulation, in part because individual genes are easier to access. The whole genome is much more complex, and much harder to study. No doubt much research into gene regulation will continue to look at the effects of local changes. But, with the advent of genomics and proteomics, as well as techniques such as that used by Vogelauer *et al.*, our view of gene regulation is becoming less restricted. ■

Shelley L. Berger is in the Molecular Genetics Program, The Wistar Institute, 3601 Spruce Street, Philadelphia, Pennsylvania 19104, USA.
e-mail: berger@wistar.upenn.edu

1. Vogelauer, M., Wu, J., Suka, N. & Grunstein, M. *Nature* **408**, 495–498 (2000).
2. Van Holde, K. E. in *Chromatin: Springer Series in Molecular Biology* (ed. Rich, A.) 111–148 (Springer, New York, 1988).
3. Bradbury, E. M. *Bioessays* **14**, 9–16 (1992).
4. Brownell, J. E. *et al. Cell* **84**, 843–851 (1996).
5. Taunton, J. *et al. Science* **272**, 408–411 (1996).
6. Grant, P. A. *et al. Genes Dev.* **11**, 1640–1650 (1997).
7. Utley, R. T. *et al. Nature* **394**, 498–502 (1998).
8. Laherty, C. D. *et al. Cell* **89**, 349–356 (1997).
9. Zhang, Y. *et al. Cell* **89**, 357–364 (1997).
10. Kadosh, D. & Struhl, K. *Cell* **89**, 365–371 (1997).
11. Nagy, L. *et al. Cell* **89**, 373–380 (1997).
12. Heinzel, T. *et al. Nature* **387**, 43–48 (1997).
13. Alland, L. *et al. Nature* **387**, 49–55 (1997).
14. Kuo, M.-H. *et al. Genes Dev.* **12**, 627–639 (1998).
15. Rundlett, S. E. *et al. Nature* **392**, 831–835 (1998).
16. Kadosh, D. & Struhl, K. *Mol. Cell. Biol.* **18**, 5121–5127 (1998).
17. Cohen, B. A. *et al. Nature Genet.* **26**, 183–186 (2000).

Chemistry

Clipping the carbon–carbon bond

Robert H. Crabtree

The bonds between two carbon atoms tend to be hard to break. But careful manipulation of the starting material can make the process remarkably easy.

Some chemical reactions are simple to state but difficult to do. One such reaction is the rapid cleavage of a carbon–carbon bond by inserting a metal atom between the two carbons. This is an important process, as it ‘activates’ the bond and allows potentially more useful molecules to be made from relatively unreactive starting materials.

A few years ago, Milstein and co-workers from the Weizmann Institute at Rehovot, Israel, successfully demonstrated this reac-

tion¹. Now, in the *Journal of the American Chemical Society*², they report a full characterization of this unusual process, which reveals that it happens so readily it even occurs at temperatures as low as -70°C . This is remarkable because previous C–C bond-breaking reactions³ have generally required temperatures of at least 100°C .

In the past, such reactions have led to valuable applications in which a small amount of the metal — acting as a catalyst —

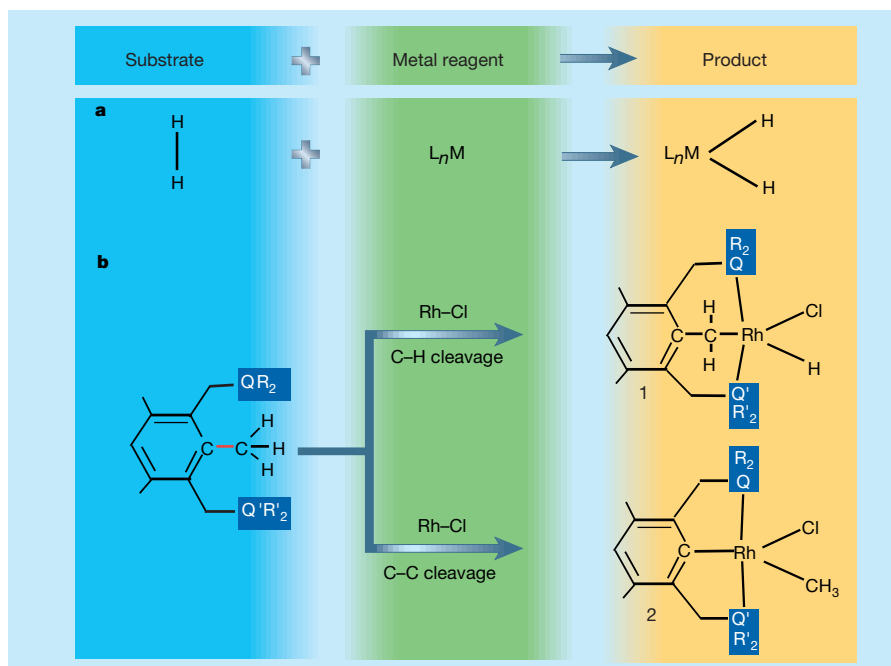
triggers the conversion of large amounts of starting materials (the substrate) into useful products. Catalysis is used in a wide variety of areas from petroleum refining to medicinal drug synthesis.

Metal catalysis occurs naturally in the enzyme reactions of biochemistry, as well as in the organic reactions of synthetic and industrial chemistry. An important goal in the chemistry of the transition metals is to understand why these metals — found in the central part of the periodic table — so effectively catalyse such a wide variety of reactions. Typical metal catalysts consist of a transition-metal atom surrounded by a set of ligands that determine the catalytic behaviour of the central metal.

In many catalytic reactions, the first step of the pathway is the insertion of a metal atom into a single bond of one of the starting materials, a reaction known as ‘oxidative addition’⁴. In one of the earliest successes in this area, molecular hydrogen, H_2 , a relatively unreactive molecule, underwent rapid oxidative addition (see Fig. 1a).

As catalysis developed, less reactive — chemically strong — bonds were tackled. For example, the carbon–hydrogen bond in an alkane (such as ethane, $\text{CH}_3\text{—CH}_3$) is particularly strong, making the molecule very stable. In fact, the bond is even less reactive than the H–H bond in molecular hydrogen, and resisted cleavage by oxidative addition for many years. In the 1970s, such reactions were found and used in catalytic dehydrogenation reactions⁵, which convert unreactive alkanes into more chemically reactive alkenes (such as ethene, $\text{CH}_2=\text{CH}_2$) by removing H_2 . In the past decade, ways have also been found to break the extremely inert fluorocarbon C–F bonds in the same way, and this has been used to convert chlorofluorocarbons into

Figure 1 Using metal complexes to break bonds. a, An early example of bond cleavage using a transition-metal catalyst. The unreactive bond in molecular hydrogen is broken when the metal (M), attached to ligands (L_n), is inserted between the two hydrogens. b, In the reaction studied by Milstein and colleagues², the C–C bond in the substrate (red) is impossible to break with conventional reagents, but it can be cleaved using the rhodium complex $[(\text{C}_2\text{H}_4)_2\text{RhCl}]_2$. Incorporating phosphorous groups ($\text{P}(\text{C}_4\text{H}_9)_2$) and/or nitrogen groups ($\text{N}(\text{C}_2\text{H}_5)_2$) at positions close to the C–C bond (Q and Q') causes the substrate to bind tightly to the metal reagent (source of the Rh–Cl fragment). If the phosphorus groups are present at both Q and Q', then one of the C–H bonds can be cleaved to give product 1, or — much harder — the C–C bond can be cleaved to yield product 2. But if a nitrogen group is introduced at Q', then only the C–C bond is broken and product 2 generated.





100 YEARS AGO

In the October issue of the *American Naturalist*, Prof. H. F. Osborn reconsiders the evidence in favour of the existence of a common ancestral stem from which have diverged dinosaurs and birds. It is argued that many of the resemblances between these groups are adaptive rather than genetic, while the apparent close correspondence in the structure of the pelvis between adult birds and the herbivorous dinosaurs (which are specialised types) is due in a considerable degree to a misinterpretation of the homology of some of their elements. Nevertheless, the resemblances between the two groups are so numerous as to justify the belief of kinship. And special importance attaches to the opinion that some sort of bipedalism was a common character of all dinosaurs, the suggestion being countenanced that certain forms, like *Stegosaurus*, have reverted from a bipedal to a quadrupedal mode of progression. From *Nature* 22 November 1900.

50 YEARS AGO

To inquire into the early stages of development of a scientific theory is no idle curiosity. Indeed, no adequate appreciation of its present state is possible if the various aspects which have gained prominence in the course of time are not traced back to their roots. This is especially true for the quantum theory of atomic systems, which presents such unfamiliar features and has led to such far-reaching conceptions as Niels Bohr's idea of complementarity. The history of quantum mechanics is, in fact, particularly instructive as an illustration of the inherent logic of the growth of scientific ideas. At the same time, it is not lacking in dramatic surprise: it forcefully reminds us that a vigorous flight of imagination can sometimes unexpectedly quicken the pace of a more methodical progress. It would be futile to try to date the birth of quantum mechanics from a particular discovery, or to relate it to a particular name... Nevertheless, it is possible to point to one inspiring idea, which gave the whole development unfailing guidance: it is Bohr's principle of correspondence, or (as he prefers to call it) correspondence "argument"... The young physicists who nowadays solve wave-equations as a matter of routine are not always taught how essential the correspondence point of view remains for the physical interpretation of the new formalism. Still less can they realize how essential an instrument it has been in the pioneering period of elaboration of this theory. From *Nature* 25 November 1950.

more environmentally friendly hydrochloro-fluorocarbons, for example^{4,6}.

Perhaps the most challenging of all such oxidative additions involves the C–C bond, not only because it is relatively strong, but also because it is always buried so deep within the molecule. The reaction studied by Milstein and colleagues² is shown in Fig. 1b. The C–C bond to be broken, shown in red, is first enticed into close proximity with the metal reagent (rhodium) by attaching groups containing phosphorus or nitrogen atoms to the substrate (at points Q and Q'). These atoms initially bind very strongly to rhodium, and in doing so draw the central ring with its target C–C bond close to the metal. Because the phosphorus atom is larger than nitrogen, the orientation of the C–H and C–C bonds in the substrate is affected. This allows the substrate containing two phosphorus atoms to undergo both the easier C–H bond cleavage to yield product 1, and the more difficult C–C cleavage to give product 2. But when the smaller nitrogen atom replaces one phosphorus, the substrate orientation is altered so that rhodium only has ready access to the C–C bond. Rapid breaking of this bond occurs and only structure 2 is produced^{3,4}.

The results of Milstein and colleagues' investigations² indicate that if the metal

reagent can be brought close enough to the C–C bond, the actual barrier to the oxidative addition is relatively low. The next step is to investigate whether this concept can be developed for use in industrial processes. One promising possibility would be a technique for rearranging the carbon skeleton of the substrate by selectively breaking its C–C bonds. This sort of chemical surgery can be achieved through catalytic cracking—widely used in the petrochemical industry—but only at temperatures of around 500 °C. In contrast, Milstein and colleagues' reaction occurs rapidly at room temperature. But it relies on the phosphorus and nitrogen groups to direct the rhodium catalyst to the desired bond. To become useful in industrial catalysis, such reactions must work without needing these directing groups. This remains a formidable challenge for the future. ■

Robert H. Crabtree is in the Department of Chemistry, Yale University, 225 Prospect Street, New Haven, Connecticut, 06520-8107, USA.

e-mail: robert.crabtree@yale.edu.

1. Gandelman, M. *et al.* *Organometallics* **16**, 3981–3986 (1997).
2. Gandelman, M., Vignalok, A., Konstantinovskii, L. & Milstein, D. *J. Am. Chem. Soc.* **122**, 9848–9849 (2000).
3. Crabtree, R. H. *Chem. Rev.* **85**, 245–269 (1985).
4. Crabtree, R. H. *The Organometallic Chemistry of the Transition Metals* 3rd edn 149–361 (Wiley, New York, 2001).
5. Jensen, C. M. *Chem. Commun.* 2443–2449 (1999).
6. Burdeniuc, J. *et al.* *Chem. Ber.* **130**, 145–154 (1997).

Evolutionary biology

Déjà vu

J. J. Bull

A long-term study of fruitflies adds to the evidence that evolution can run backwards. To what extent the genetic underpinnings revert to the original is unclear.

We tend to think of evolutionary change as resulting in something different or new. But evolution going backwards, in which an organism reverts to a previous state, has its own possibilities and puzzles. Teotónio and Rose¹ (page 463 of this issue) show that, in captive populations of the fruitfly *Drosophila melanogaster*, the effects of 100–200 generations of adaptation to a particular environment can be undone in as little as 20 generations when the flies are returned to their original environment. A few traits did not revert, however, even when the flies had enough genetic variation to go all the way back. These cases shed light on why evolution does not necessarily reverse.

Evolution undoing previous evolution has often been thought to be improbable. But one must distinguish between molecular genetic and phenotypic reversibility. In the first instance, the question is whether the original genetic changes have been undone, recreating the ancestral genotype. In the second, the question is whether the organism's morphological, physiological or behaviour-

al traits have reverted to an ancestral state (for example, the re-evolution of wings in a flightless insect), without regard to the genetic basis of that reversion. As in most earlier work on reversibility, Teotónio and Rose look at the evolution of phenotypic traits.

The common view that phenotypic evolution is largely irreversible is codified in Dollo's law, dating from the late 1800s. A broad interpretation of this law is simply that evolution cannot go back to recreate an organism in exact detail. More narrowly, and sometimes recognized as Abel's law or Meyrick's law, it has been interpreted to mean that lost structures (wings, say) are not regained^{2,3}. The narrow sense of Dollo's law certainly has its exceptions. Darwin recognized that domesticated organisms sometimes regain ancestral features when returned to the wild. Many lab experiments also show that when artificial selection on a trait is relaxed, that trait often reverts fully or partially to the ancestral condition⁴.

A striking case of reversibility is seen in the live, attenuated virus of the Sabin polio-

Materials science

Pore characterization

Porous materials are extremely useful in catalysis and selective molecular filtration, and can even serve as miniature laboratories for conducting chemistry with just a few molecules at a time.

Most natural porous minerals, such as zeolites, are laced with pores whose width is no greater than the size of small molecules (less than 2 nm). But materials scientists and chemical engineers want to make synthetic 'mesoporous' solids with larger pores (2–50 nm) that are selective for larger molecules.

Although much effort has gone into finding new ways of making such mesoporous solids, the atomic-scale structure of the pores is not well known and so refining their structure is a matter of trial-

and-error. This is due partly to the difficulty in obtaining large single crystals for conventional diffraction studies using X-rays or neutrons. Elsewhere in this issue (*Nature* **408**, 449–453; 2000), Galen Stucky and colleagues report a method that uses high-resolution electron microscopy to obtain images that can be mathematically treated to give a full three-dimensional (3D) structure of mesoporous materials.

Electron microscopy uses electrostatic lenses to form a magnified image of the sample (by recombining the diffracted electron beams). This has the advantage that it can be carried out on very small areas, because the focused electron beam is only a few micrometres in diameter. The technique does not depend on the actual resolution of

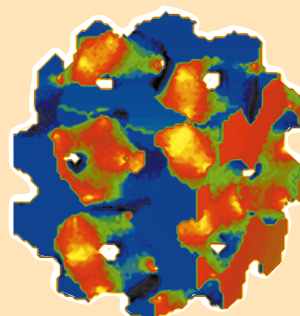
the images, but provides more complete information about the structure. So, compared with conventional methods for structural analysis of crystalline and amorphous materials, this approach is generally applicable to soft materials that have disorder at the atomic length scale, but order at the mesoscopic scale.

The picture here shows the 3D structure of a mesoporous silica material obtained by Stucky and colleagues. This image provides information about the sizes and shape of the pores at the nanoscale level, as well as their connectivity. Further analysis reveals an unusual bimodal structure, in which there are different sized micro- and mesopores.

What makes this imaging method so attractive is that, compared with

X-ray diffraction, it provides full structural information without using pre-assumed models or parameters. Moreover, it can reveal high-resolution details of pore structure at different scales in porous composites. The technique could be used to characterize the detailed structure of a wide range of composite materials.

Vincent Dusastre



virus vaccine. This virus often reverts to a more virulent form within the vaccinated person; such reversion does not harm the person concerned but can cause polio when transmitted to others who have not been vaccinated or previously infected⁵.

Teotónio and Rose conducted a large experiment using caged populations of *Drosophila* that were originally maintained for other purposes. For 25 years, one population has been kept under standard rearing conditions (the 'standard' line). Twenty years ago, a subset of the standard flies was used to start another population that was then selected for late-life fecundity (this line was originally used to test evolutionary theories of senescence). Eleven years ago, flies from the standard line were used to found two additional populations, one being selected for resistance to starvation, the other for intermediate generation time. Finally, eight years ago, a fourth population was started from the standard line, selecting for flies that developed into adults especially quickly.

About three years ago, flies from these four lines were returned to the original standard conditions. Since then, eight different phenotypic traits have been monitored across five replicates of each line: male and female development time; male and female resistance to starvation; early fecundity at high and low population densities; female dry body weight; and female lipid content. Teotónio and Rose observed three different patterns of reversibility: complete, partial and none. Most traits reversed completely, although the time taken for this varied from 20 to 50 generations. Partial reversals were not merely slower versions of complete

reversals, because their initial rate of reversal was rapid but then reached a plateau.

Why did some traits — such as fecundity at high population densities — fail to reverse at all or to completely attain the ancestral condition? The limitation could be genetic — that is, genetic variation was absent or exhausted by selection. Or epistasis could have blocked the return. Epistasis is an interaction in the effects of different mutations, such that the consequence of one mutation changes depending on whether another mutation is present or absent.

An example is seen in bacteria that have become resistant to antibiotics. Although mutations that give rise to drug resistance are, by themselves, generally deleterious in the absence of antibiotic, the ending of antibiotic treatment does not ensure the evolutionary reversal to antibiotic sensitivity. This is because drug-resistance mutations often impair essential cellular functions and so are followed by other, compensatory mutations that restore those functions. In the absence of antibiotics, the antibiotic-resistance gene and the compensatory mutations are each singly more harmful than they are in combination. So reverse evolution is blocked because neither gene can undergo reversal on its own⁶.

To explore similar possibilities in the fruitfly, Teotónio and Rose created hybrid populations between the selected and standard lines. These hybrid populations were genetically variable, and so should have reverted more easily because they lacked many types of genetic barriers to reversion. Correcting for the fact that the hybrid populations started their return from intermediate values, the replicates behaved essentially

as the evolved populations did, suggesting that genetic limitations did not explain the lack of reversion. Teotónio and Rose speculate that the form of selection was instead altered by the original evolution — that is, for some traits the nature of selection depended in subtle ways on the genetic background of the population, which changed over the course of the experiment.

But is phenotypic reversal accompanied by genetic reversal? Genetic reversal seems less likely than phenotypic reversal, because there are often several genetic routes to an adaptation, and because mutational biases and epistasis may effectively preclude a reversal back down the original pathway. Genetic reversions have been observed in microbes^{7,8}, but resolution of this question in *Drosophila* awaits further work. Nonetheless, Teotónio and Rose's study adds to a growing body of work on adaptation, in which natural or semi-natural replicates tend to reveal that natural selection has a powerful influence on phenotype. Their work suggests that there is essentially no intrinsic barrier opposing the return to formerly adaptive states, and that reversals of many types of phenotypic traits can occur. ■

J. J. Bull is in the Section of Integrative Biology and Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas 78712, USA.
e-mail: bull@bull.biosci.utexas.edu

1. Teotónio, H. & Rose, M. R. *Nature* **408**, 463–466 (2000).
2. Gould, S. J. *J. Hist. Biol.* **3**, 189–212 (1970).
3. Minkoff, E. C. *Evolutionary Biology* (Addison Wesley, Reading, MA, 1983).
4. Hill, W. G. & Caballero, A. *Annu. Rev. Ecol. Syst.* **23**, 287–310 (1992).
5. Guillot, S. *et al. Vaccine* **12**, 503–507 (1994).
6. Schrag, S. J. *et al. Proc. R. Soc. Lond. B* **264**, 1287–1291 (1997).
7. Crill, W. D. *et al. Genetics* **154**, 27–37 (2000).
8. Levin, B. R. *et al. Genetics* **154**, 985–997 (2000).

Developmental biology

Control by combinatorial codes

Arjumand Ghazi and K. VijayRaghavan

Studies in fruitflies support the idea that regulatory regions of genes control development by acting as molecular 'computers', calculating cell fate according to the combined effects of several signalling pathways.

Multicellular organisms are made up of a huge number of cell types, specialized to perform particular functions. During development, the fate of a cell is determined by outside signals that ultimately lead to specific changes in gene expression inside the cell. But, compared to the number of cell types, the number of signalling pathways is tiny, so each must be quite general. How, then, are these signals translated into specific patterns of gene expression? Three groups, writing in *Cell*^{1–3}, address this problem in different developmental contexts. Their results strengthen a view of gene regulation⁴ as the deciphering of a combinatorial code, in which the quality and quantity of signals are directly sensed and computed by regulatory elements in the target gene.

Developmental decisions can be seen as a set of logical operations that integrate the effects of short- and long-range signals through the action of transcription factors⁴ (proteins that help to transcribe genes into

messenger RNAs). Take, for example, the case of muscle formation in the fruitfly *Drosophila melanogaster*, where several types of signalling pathway result in the gene *even-skipped* being turned on in just one cell of a cluster of mesodermal cells (Fig. 1a). Mesoderm is the tissue from which muscle develops, and cells that express *even-skipped* go on to form particular types of heart and body-wall muscles. What is the molecular 'computer' that calculates the effects of the signalling pathways on the *even-skipped* gene? Halfon *et al.*¹ propose an answer, in the shape of an 'enhancer' — a regulatory region of the gene.

The authors¹ find that this particular enhancer region is needed to ensure the expression of *even-skipped* in the requisite cell from a cluster of mesodermal cells. Within the enhancer region, the authors identify sites that bind regulatory factors positioned near the ends of receptor-tyrosine-kinase signalling pathways. They also find regions

that bind to factors from the Wingless and Decapentaplegic signalling systems. Finally, they find binding sites for Twist, a protein specific to undifferentiated mesodermal cells, and for Tinman, a marker of dorsal mesodermal cells.

By altering these binding sites, or by changing the combination of regulatory factors that are expressed, Halfon *et al.* show that all of these signalling pathways, and all of the binding sites, are needed to achieve proper expression of *even-skipped* (Fig. 1b). Here, then, the combinatorial code involves the Wingless, Decapentaplegic and receptor-tyrosine-kinase signalling pathways, as well as the tissue-specific expression of Twist and Tinman. The code is deciphered by the enhancer of the *even-skipped* gene.

In the *Drosophila* eye, cone cells, pigment cells and some types of photoreceptor cell develop from a common pool of precursor cells in response to different signals. The precursor cells use receptor-tyrosine-kinase pathways, sometimes in the context of signalling through the Notch pathway, to make developmental decisions. Flores *et al.*² and Xu *et al.*³ show that the enhancer regions of target genes are again at the heart of these decisions.

Cone cells form from a cluster of 'equivalent' precursor cells that are responsive to the *Drosophila* epidermal-growth-factor receptor (DER; a receptor tyrosine kinase), and that express the transcription factor Lozenge. Signalling through the Notch pathway in some of these cells results in the expression of the transcription factor D-Pax2, and a fate as a cone cell. Flores *et al.*² show that the combinatorial code for expression of D-Pax2 is deciphered by the enhancer region of the D-Pax2 gene, and involves Lozenge and proteins from the Notch and DER pathways.

Finally, Xu *et al.*³ study the specialization of so-called R7 photoreceptor cells, which involves changes in the levels of the *prospero* gene in a group of the precursor cells. The authors identify an enhancer in the *prospero* gene that ensures low levels of *prospero* expression, a feature of the early stages of R7-cell specialization. This enhancer also responds to DER signalling, and has binding sites for Lozenge. In other words, the combinatorial code that produces these early, low levels of *prospero* expression is deciphered at the *prospero* enhancer region.

Of the cells that express Lozenge, respond to DER and express low levels of *prospero*, those that pick up the Notch signal activate expression of D-Pax2 and become cone cells. However, those cells in which the Sevenless protein, a receptor tyrosine kinase, is activated, but in which Notch signalling is absent, express high levels of *prospero* — a feature of the late stages of R7-cell specialization — and become R7 cells³. Flores *et al.*² confirm this: according to this model, Notch

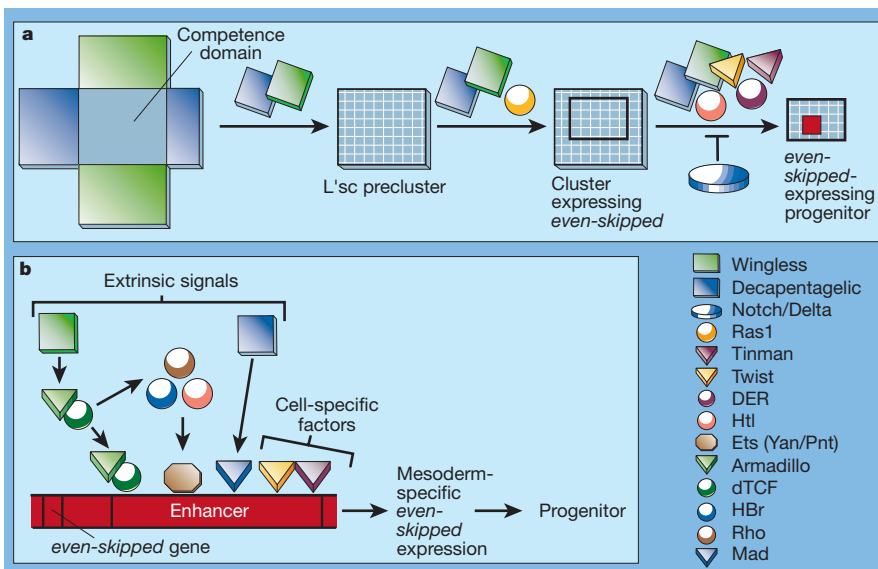


Figure 1 Gene regulation by combinatorial logic. Three new papers^{1–3} provide further evidence for the 'computational' role of gene enhancer regions. The developmental system studied by Halfon *et al.*¹ is shown here. a, An early stage in the development of heart and body-wall muscle in *Drosophila* is the production of a cell that expresses the *even-skipped* gene. From left, stripes of Wingless (vertical) and Decapentaplegic (horizontal) signals along the epidermis (the outer layers of an embryo) act on underlying mesodermal cells expressing the proteins Twist and Tinman, to define a cluster of 'competent' cells expressing the transcription factor L'sc. Signalling through the Wingless, Decapentaplegic and receptor-tyrosine-kinase pathways (the latter involving the proteins DER and Htl) eventually produces just one *even-skipped*-expressing cell. Modified from ref. 5. b, The enhancer as a computer. Proteins downstream of extrinsic signals (Wingless, Decapentaplegic and receptor-tyrosine-kinase signals combine with intrinsic factors (Twist and Tinman) to regulate the expression of *even-skipped*. (Htl, Hbr and Rho are components of receptor-tyrosine-kinase signalling pathways.)

signalling in 'presumptive' R7 cells should convert them into cone cells, and this is what happens.

So it seems that the transcription factors that define domains in a tissue — such as Twist in undifferentiated mesoderm, or Lozenge in some precursor cells in the eye — are continually required in a given cell lineage. Some signalling pathways, such as those involving Wingless and Decapentaplegic, may be similarly required. They provide a background, or context, that is probably continuously sensed by the target gene. The 'final' choice of fate is determined by more general inductive signals, such as regulated receptor-tyrosine-kinase signalling acting within this background. The combinatorial logic allows the repeated use of a small number of elements to give a unique output. This way of controlling gene expression requires that all types of signal act directly at the target.

The next level of complexity is presumably provided by the existence of autoregulatory and feedback circuits that are kicked off by the protein product of the target gene. It will be interesting to find out how stable such circuits are. And a cell's response to

signals is not limited to modifying gene expression. It will also modify proteins, for example, or move proteins around the cell, and these processes will add further layers of complexity.

Earlier analysis of gene regulation in sea urchins⁴ showed the possibility of modelling gene regulation as logic circuits, and revealed the computational power of enhancers. The latest three papers^{1–3} provide much-needed data to allow further development of such models. And as more gene circuits are unravelled, we may well find that the formal logic underlying their regulation has much to contribute to our understanding of other complex circuits, such as the neuronal circuits seen in the brain. ■

Arjumand Ghazi and K. VijayRaghavan are at the National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India.

e-mail: vijay@ncbs.res.in

1. Halfon, M. S. *et al.* *Cell* **103**, 63–74 (2000).
2. Flores, G. V. *et al.* *Cell* **103**, 75–85 (2000).
3. Xu, C., Kauffmann, R. C., Zhang, J., Kladny, S. & Carthew, R. W. *Cell* **103**, 86–94 (2000).
4. Yuh, C.-H., Bolouri, H. & Davidson, E. H. *Science* **279**, 1896–1902 (1998).
5. Carmena, A. *et al.* *Genes Dev.* **12**, 3910–3922 (1998).

was administered to autoimmune diabetic 'NOD' mice⁴.

There are several clever features to this approach that contributed to the authors' success. The native insulin gene encodes proinsulin, which is a protein consisting of two chains, the A and B chains, linked by a 35-amino-acid sequence, the 'C peptide'. To produce insulin, proinsulin must be processed by proteolytic (protein-cleaving) enzymes in the pancreatic β -cells to remove the C peptide⁵. This releases the A and B chains, which are then connected and folded to create native insulin⁵.

The proteolytic processing steps are complex, and rely on cleavage enzymes and other machinery⁶. This has been a significant obstacle in the past, because most tissues do not express the proper cleavage enzymes and other factors needed to process proinsulin. Attempts to overcome this problem by inserting alternative cleavage sites in the proinsulin molecule, which can be recognized by liver-specific proteolytic enzymes, have been only partially successful. Lee *et al.*³ have addressed this problem by replacing the 35 C-peptide residues with a short, seven-amino-acid peptide, avoiding the need for proteolytic processing. Whereas proinsulin has only 1–2% of native insulin's biological activity, Lee *et al.*'s single-chain insulin analogue was 20–40% as active as native insulin.

Normally, glucose is the main molecule that induces the secretion of insulin from the pancreas. An ideal gene-therapy system would require insulin secretion to be similarly responsive to changes in blood glucose levels. Lee *et al.*³ achieved this by placing the gene encoding the single-chain insulin analogue under the control of the glucose-responsive L-type pyruvate kinase promoter. Surprisingly, this seems to have worked reasonably well. Ready-made insulin is normally stored in pancreatic β -cells and is secreted by a process of stimulated 'exocytosis', which allows rapid responses to molecules such as glucose. With the viral construct, changes in glucose levels modulate expression of the single-chain-insulin gene. By definition, this is a sluggish process compared with stimulated exocytosis, and one would expect that secretion of the insulin analogue in response to glucose would be substantially delayed. This was in fact seen in Lee *et al.*'s study. Remarkably, however, the changes in blood glucose levels were not that different between healthy control and virus-treated animals.

The virus-treated animals were able to maintain reasonably normal fasting (between-meal) blood glucose levels over the 8-month study period. Although important, this degree of 'normoglycaemia' might reflect unregulated secretion of the single-chain insulin. It is well known that basal levels of insulin secretion have a major impact on fasting glucose levels⁷. However,

Diabetes

Gene therapy for rats and mice

Jerrold M. Olefsky

Delivery of an insulin-encoding gene into diabetic rats and mice has helped them to regulate their blood glucose levels. But there are still obstacles to overcome before this approach can be used in humans.

Type I diabetes is an autoimmune disease that begins in childhood or early adulthood, and eventually causes complete destruction of the insulin-secreting β -cells in the pancreas¹. The upshot is that patients with type I diabetes produce absolutely no insulin. So attempts to treat this disease have centred on ways to administer externally produced insulin — by subcutaneous injection, external pumps or nasal sprays, to name but a few. The goal is to match insulin delivery to changes in blood sugar levels so as to achieve normal glucose levels. But all of these methods suffer from inconvenience, interference with lifestyle and, most important, the near impossibility of matching insulin administration to the body's minute-to-minute needs. As a result, much effort is going into developing alternative forms of treatment that bypass the need for external insulin, such as the transplantation of pancreatic islet cells², the development of replenishable insulin-secreting cell lines for implantation, and, more recently, gene therapy.

On page 483 of this issue³, Lee and colleagues describe an advance that may pave

the way to the use of insulin gene therapy in patients with type I diabetes. Using genetic engineering, the authors have produced an insulin analogue that retains an impressive 20–40% of the activity of native insulin. They inserted the DNA encoding this insulin analogue into a modified virus for delivery into diabetic animals. They then placed the entire DNA construct under the control of the promoter region of the L-type pyruvate kinase gene found in liver cells. This promoter is a regulatory element that normally controls the expression of the L-type pyruvate kinase gene in response to glucose.

The authors injected the resulting virus intravenously into rats and mice, and found that the viral genome became incorporated exclusively in liver cells. The animals were able to secrete the biologically active insulin analogue from their livers into their bloodstreams for up to 8 months. When rats were made diabetic by administration of the β -cell toxin streptozotocin, and then treated with the viral construct, they maintained normal blood glucose levels throughout the entire 8-month study period. Similar results were obtained³ when the virus

constant basal insulin levels are inadequate to control glucose concentrations after eating⁸, and some method for timing insulin delivery to glucose ingestion is needed. This is why it is interesting that the virus-treated animals controlled blood glucose levels so well after eating, despite the delayed kinetics of insulin secretion.

But rodents are quite different from humans with respect to maintaining glucose levels, and extending these results to human physiology may prove a challenge. Rodent livers have much higher basal rates of glucose production than human livers^{9,10}. So small effects of insulin on the liver may be able to control post-eating glucose levels in rodents, but may be less effective in humans. In addition, given that the viral construct integrated directly into the chromosomes of the liver cells, these cells may be exposed to abnormally high local levels of insulin, which are not reflected in the measurements of single-chain-insulin concentrations in the blood. One would also expect that the delayed and prolonged insulin secretion in the virus-treated animals would lead to hypoglycaemia (abnormally low glucose levels) several hours after glucose ingestion. Indeed, post-digestion blood glucose levels were lower in the virus-treated animals than in controls.

Whether Lee *et al.*'s approach³ can be applied to humans remains to be determined. One of the biggest obstacles to anti-diabetes therapy is the hypoglycaemia that often accompanies efforts to maintain tight glucose control. Administration of too much insulin through an irreversible gene-delivery system is of concern. And an individual's overall insulin secretion changes with diet, body weight, age, growth status, and so on. So a gene-therapy approach must be able to modulate insulin delivery in response to changing needs over time. Placing insulin secretion under the control of glucose is nature's way of overcoming this problem, and Lee *et al.*'s use of a glucose-responsive promoter in the viral vector is a definite step forward. Despite these issues, the paper represents a good example of how basic research can be applied to problems of clinical significance.

*Jerrold M. Olefsky is in the Department of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0673, and the San Diego VA HealthCare System, San Diego, California 92161, USA.
e-mail: jolefsky@ucsd.edu*

1. Eisenberth, G. S., Ziegler, A. G. & Colman, P. A. in *Joslin's Diabetes Mellitus* 13th edn (eds Kahn, C. R. & Weir, G. C.) 216–239 (Lea & Febiger, Philadelphia, 1994).
2. Shapiro, A. M. *et al.* *N. Engl. J. Med.* **343**, 230–238 (2000).
3. Lee, H. C., Kim, S.-J., Kim, K.-S., Shin, H.-C. & Yoon, J.-W. *Nature* **408**, 483–488 (2000).
4. Pozzilli, P. *et al.* *Immunol. Today* **14**, 193 (1993).
5. Rhodes, C. J. in *Diabetes Mellitus: A Fundamental and Clinical Text* 2nd edn (eds LeRoith, D., Taylor, S. I. & Olefsky, J. M.) 20–37 (Lippincott Williams & Wilkins, Philadelphia, 2000).

6. Shoelson, S. E. & Halban, P. A. in *Joslin's Diabetes Mellitus* 13th edn (eds Kahn, C. R. & Weir, G. C.) 29–55 (Lea & Febiger, Philadelphia, 1994).
7. Rosenzweig, J. L. in *Joslin's Diabetes Mellitus* 13th edn (eds Kahn, C. R. & Weir, G. C.) 460–488 (Lea & Febiger, Philadelphia, 1994).

8. Brink, S. J. in *Diabetes Mellitus: A Fundamental and Clinical Text* 2nd edn (eds LeRoith, D., Taylor, S. I. & Olefsky, J. M.) 488–499 (Lippincott Williams & Wilkins, Philadelphia, 2000).
9. Kolterman, O. G. *et al.* *J. Clin. Invest.* **68**, 957–969 (1981).
10. Miles, P. D. G. *Diabetes* **46**, 1678–1683 (1997).

Molecular electronics

A dual-action material

Fernando Palacio and Joel S. Miller

In the drive for smaller electronic components, chemists are thinking on a molecular scale. By combining two simple molecules, a hybrid has been produced that is both magnetic and an electrical conductor.

As we approach the physical limits of conventional silicon-based electronics, there is a clear need for entirely new types of materials that can deliver smaller and smaller devices. A strong candidate in this field is molecular electronics, which uses assemblies of individual molecules to mimic larger, conventional structures such as switches or semiconductors. To be effective, full control over the composition, size and function of these molecules is needed.

Over the past 40 years, researchers have created some of the basic building blocks needed for molecular electronics, including metals, semiconductors, superconductors and magnets, as well as the so-called single-molecule magnets. On page 447 of this issue Coronado *et al.*¹ describe a hybrid material that opens a new frontier in molecular electronics. They have created a molecular material that can behave simultaneously as a magnet and an electrical conductor. Whereas conducting magnets, such as nickel or iron, are common among metals and alloys, this is the first reported example in molecular materials. Because the electronic

peculiarities of individual molecules are different from those of bulk metals, a conducting molecular magnet is likely to have unexpected properties.

Materials whose properties are based on their component molecules are more versatile than those whose properties derive from their component atoms. Thus the bulk properties of molecules — whether optical, magnetic or electrical — can be controlled using conventional syntheses, such as those used in the pharmaceutical industry. This means that they are tuneable and can more readily be tailored in response to the changing demands of technology. By combining an organic conductor and a magnetic complex, Coronado *et al.* have now introduced the possibility of materials with multiple functions.

The ability of small organic molecules to exhibit metal-like electrical conductivity was first shown in 1965 in an electron-transfer salt known as TCNQ². These studies showed that metal-like conductivity as well as metal-like optical properties could be seen in soluble organic materials. This work led to the development of electrically conducting

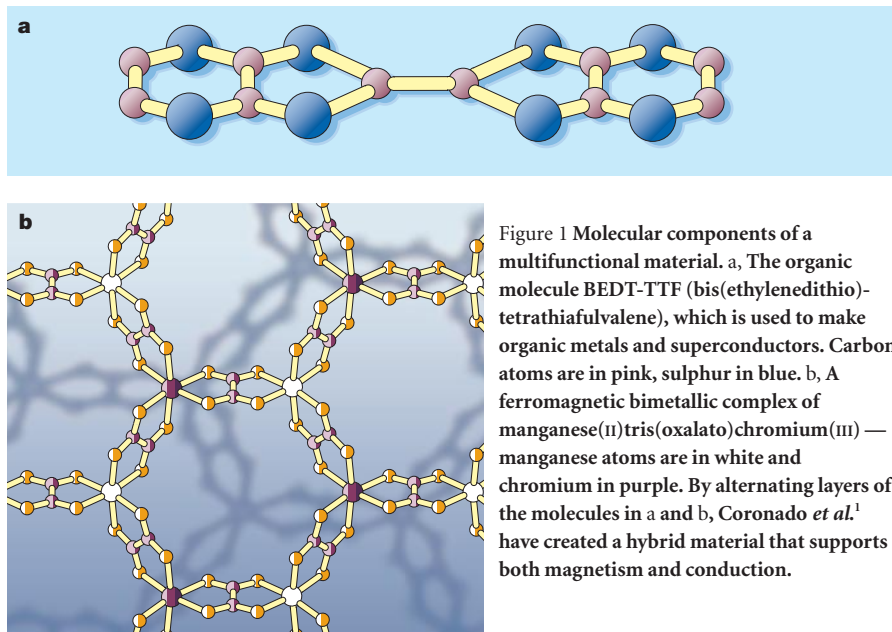


Figure 1 Molecular components of a multifunctional material. a, The organic molecule BEDT-TTF (bis(ethylenedithio)-tetrathiafulvalene), which is used to make organic metals and superconductors. Carbon atoms are in pink, sulphur in blue. b, A ferromagnetic bimetallic complex of manganese(II)tris(oxalato)chromium(III) — manganese atoms are in white and chromium in purple. By alternating layers of the molecules in a and b, Coronado *et al.*¹ have created a hybrid material that supports both magnetism and conduction.

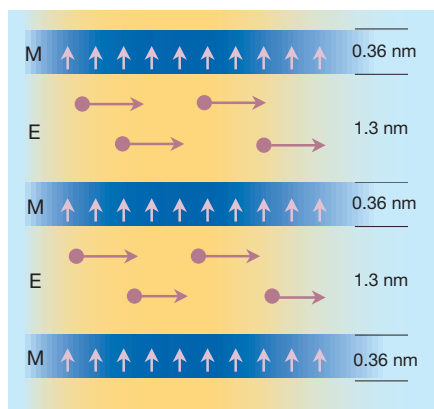


Figure 2 The molecule-based multilayer structure created by Coronado *et al.*¹, in which ferromagnetic layers (M) alternate with metal-like conducting ones (E).

polymers — essentially plastic conductors — for which Alan MacDiarmid, Alan Heeger and Hideki Shirakawa received this year's Nobel Prize for Chemistry. In addition to conducting polymers, other charge-transfer salts were discovered and studied. Of these, the molecule [TTF][TCNQ], which led to the first examples of organic superconductors, is probably the best known. Several organic superconductors³ and organic metals were made from a related molecule called BEDT-TTF (Fig. 1a). Coronado *et al.* make use of the conducting properties of this molecule in their new material.

Ferromagnetism is the parallel alignment of all the magnetic moments in a material, whether atomic or molecular, induced by applying a weak external magnetic field. This leads to a spontaneous magnetization, which may remain once the external field is

removed, as is found in the magnets used on doors, or it may disappear, as in modern current-transformers. Molecule-based ferromagnetism was first reported in 1972 for an iron chloride coordination compound^{4,5}. Then in 1986 ferromagnetism was discovered in an organic-based material⁶. These materials became ferromagnets at extremely low temperatures (below 5 K) and were soluble in conventional organic solvents. In 1992 Okawa and co-workers surprised researchers when they reported a layered bimetallic compound that behaved as a two-dimensional ferromagnet⁷. Although insoluble, these magnets are crystalline and their molecular components self-assemble into a layered structure in aqueous solutions.

Coronado *et al.*¹ created their material using controlled self-assembly of alternating single layers of a bimetallic ferromagnet (Fig. 1b) and BEDT-TTF molecules. Precise control of the assembly is possible because the BEDT-TTF layers are positively charged and the magnetic layers are negatively charged. The alternating monolayers of BEDT-TTF and the ferromagnet are just 1.3 and 0.36 nm thick, respectively (Fig. 2). So the magnetic layers develop bulk ferromagnetic ordering, but this remains independent of the current flowing in the organic layers. The authors find no evidence that the ferromagnetic and conducting subsystems interact with each other, except when an external magnetic field is applied perpendicular to the layers.

The synthesis of molecular materials that can deliver technologically important physical or chemical properties, or a combination of these properties, is now a major goal for chemists. Coronado *et al.*'s results¹ show that self-assembly allows multifunctional mat-

Daedalus

David Jones

David Jones, author of the Daedalus column, is indisposed.

erials to be made while keeping precise control over their composition and structure at the nanoscale. In the past few years we have witnessed the birth of 'spintronics' — electronic devices that are based on interactions between the 'spins' of electrons rather than their charge. A popular example is a so-called spin valve. This consists of two ferromagnetic layers separated by a weaker diamagnetic layer, which can be either an insulator or a metallic conductor only a few atomic layers thick. The ferromagnetic layers communicate with each other through electron-tunnelling effects or through the mobile electrons in the metallic layer. The material reported by Coronado's group suggests that molecule-based materials with these properties are on the horizon. This could make these devices easier to process and lead to a substantial reduction in production costs.

In conventional metallic ferromagnets, the mobile electrons play a crucial role in both the magnetic interactions and the conductivity. But in Coronado *et al.*'s system, the conducting electrons in the organic layer do not appear to interact with the magnetic moments of the ferromagnetic layer. This unique feature, which is only possible because of the molecular nature of the system, may yet yield unforeseen physical behaviour. It will also be desirable to develop hybrid molecular materials in which the conducting and magnetic subsystems do interact with each other — these could then be used to develop new electronic devices that operate at the nanoscale. It is easy to imagine other hybrid materials that combine magnetism with nonlinear optical properties, or ferromagnetism with superconductivity. The latter will be useful for exploring the interplay between superconductivity and magnetism, which, like oil and water, usually do not mix.

Fernando Palacio is at the Instituto de Ciencia de Materiales de Aragón, CSIC, Universidad de Zaragoza, 50009 Zaragoza, Spain.

e-mail: palacio@posta.unizar.es

Joel S. Miller is in the Department of Chemistry, University of Utah, Salt Lake City, Utah 84112-0850, USA.

e-mail: jsmiller@chemistry.utah.edu

1. Coronado, E., Galán-Mascarós, J. R., Gómez-García, C. J. & Laukhin, V. *Nature* **408**, 447–449 (2000).
2. Melby, L. R. *Can. J. Chem.* **43**, 1448–1454 (1965).
3. Williams, J. M. *et al. Science* **252**, 1501–1508 (1991).
4. Wickman, H. H. *J. Chem. Phys.* **56**, 976–982 (1972).
5. DeFotis, G. C., Palacio, F., O'Connors, C. J., Bhatia, S. N. & Carlin, R. L. *J. Am. Chem. Soc.* **99**, 8314–8315 (1977).
6. Miller, J. S. *et al. J. Chem. Soc. Chem. Commun.* 1026–1028 (1986).
7. Tamaki, H. *et al. J. Am. Chem. Soc.* **114**, 6974–6975 (1992).

Anthropology

Lovely grub

Western gourmets are happy to tuck into a bowl of mussels or get to grips with a lobster, but a plate of earthworms or ants would probably be politely declined. The example of Amazonian Amerindians shows that these fussy spots are passing up a nutritious and ecologically friendly food source.

From fieldwork and surveys of the literature, Maurizio Paoletti and colleagues found that invertebrates form an important part of the diets of 32 Amazonian ethnic groups (*Proc. R. Soc. Lond. B* **267**, 2247–2252; 2000). The most widely consumed species are those that feed on leaves or leaf litter, including leaf-cutter ants,

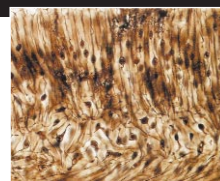
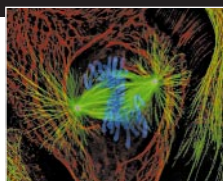


termites, caterpillars (including the cassava hornworm, pictured) and large earthworms. The last, when smoked, are a delicacy to the Ye'Kuana people of Venezuela — who even 'farm' the worms by introducing them into worm-poor patches of ground and reharvesting them.

Leaves are the most

abundant source of plant matter in the Amazon. So the animals that feed on them are a highly efficient and sustainable source of nutrients. This is a good example, say Paoletti *et al.*, of how indigenous peoples support themselves from natural resources without causing ecological damage.

John Whitfield

The Olympus and *Nature* competition 2000

A 1, 2, 3 in light microscopy

Graham A. Dunn

The standard of entries in the Olympus and *Nature* Light Microscopy Competition was extremely high. But there could be only one winner: a stunning image of a dividing newt lung cell.

Light microscopy is undergoing its greatest upheaval since Ernst Abbe radically improved the design of microscopes over a century ago, most notably by introducing the condenser, which greatly enhanced

the quality of illumination. Now, experienced microscopists, confronted by one of the new-fangled, computer-assisted microscopes, are often left wondering which knob to twiddle first. Nevertheless, the resplen-

dent array of entries in the Olympus and *Nature* Light Microscopy Competition, announced earlier this year on 6 April with a closing date of 13 October, leaves little doubt that a new breed of micro-

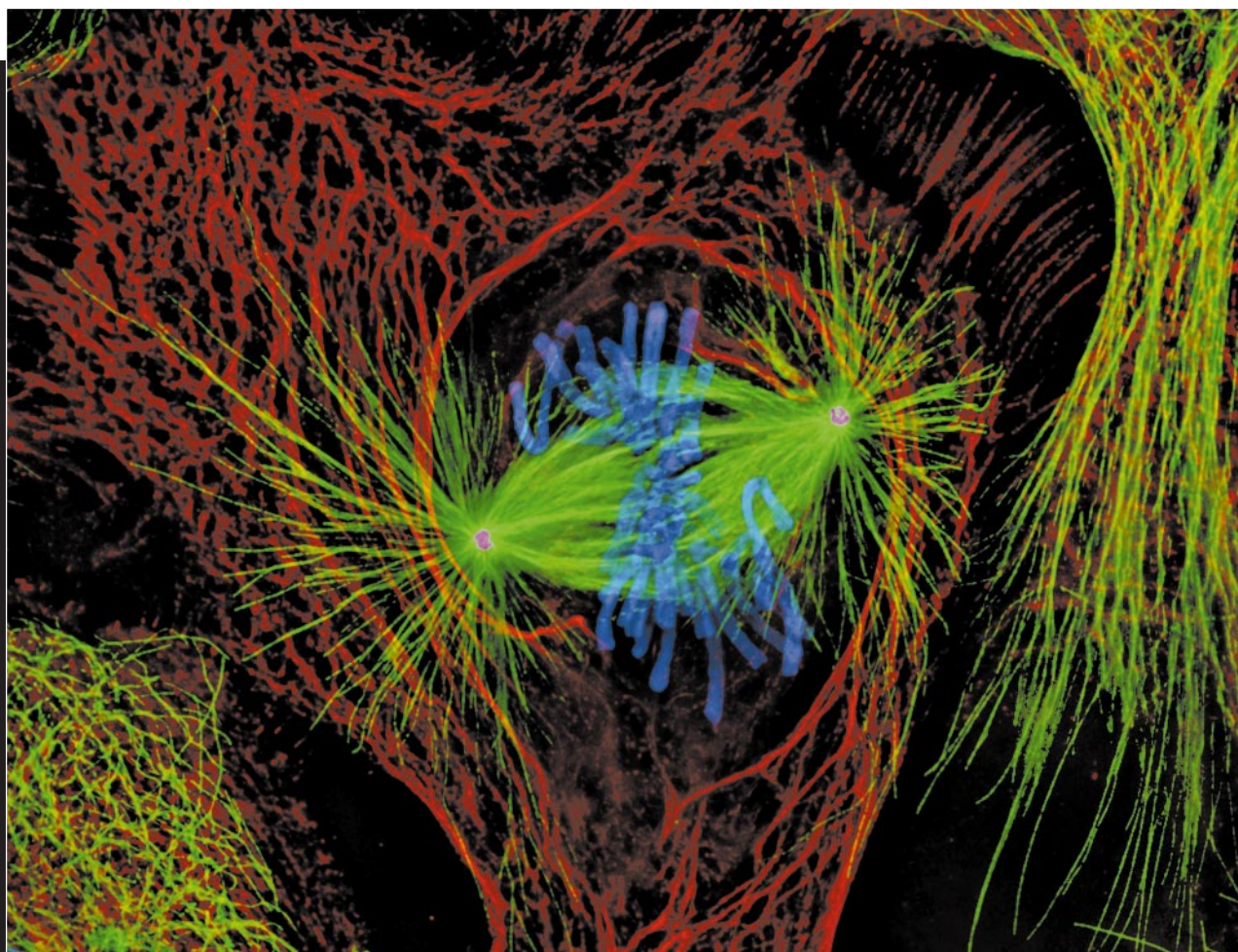


Figure 1 First prize, Alexey Khodjakov. This image of a newt lung cell shows the metaphase cell stained for centrosomes (magenta), microtubules (green), chromosomes (blue) and intermediate filaments (red). It is a maximal intensity projection through the entire cell volume. Technical details. Custom-modified TE200 microscope with 60×1.4 na objective, piezo-positioner, filter wheel and Hamamatsu Orca II camera. Magnification, $\times 1,260$.

scopists has arisen and has met the challenge heroically.

Gauging the skill shown by the competitors proved to be a challenge in itself, in that the equipment used for the entries ranged from modest student microscopes to complex and sometimes highly modified laboratory installations. But the judges had little need of lengthy deliberation to select a stunning image of a dividing newt lung cell (Fig. 1) for the star prize. It was taken by Alexey Khodjakov, a research scientist at the Wadsworth Center in Albany, New York. This technical *tour de force* shows the major structural components of the mitotic spindle that is central to the process of cell division. Centrosomes (magenta) serve as nucleating centres from which microtubules (green) grow outward. The growing ends of microtubules interact with kinetochores on the chromosomes (blue), and the whole apparatus is kept in place and separated from the rest of the cellular organelles by a cage of keratin filaments (red). As a fitting reward for this effort, Dr Khodjakov carries off the latest in compact digital camera technology: an Olympus Camedia C-2500L camera.

The second prize of a Camedia C-2000 Zoom camera goes to Eric Cho of The Chinese University, Hong Kong. He used a Zeiss Axioskop to obtain a strikingly beautiful immunofluorescence image of a hamster retina (Fig. 2), with axons shown in red and the processes of astrocytes in blue. This image reveals the intimate relationships between the axons of the retinal ganglion cells, which carry light signals to the brain, and the astrocytes and blood vessels, which play important supporting roles. For all the sophistication of the latest digital image sensors, they still have some way to go before they can match the resolving power and remarkable image processing packed into the vertebrate retina.

An attractive, traditional image of a most unusual specimen carried off the third prize of an Olympus Camedia C-920 Zoom camera. Jason Hilton, an NERC research fellow in palaeobotany and evolutionary plant biology at the University of Cardiff, used an Olympus BX 40 microscope and CCD camera to take this transmitted-light micrograph of a permineralized fossil seed nearly 300 million years old (Fig. 3). This specimen, prepared around a hundred years ago, reveals uncanny cellular detail that bears comparison to that seen in preparations from living plants.

All entries were of a consistently high standard. Thanks from myself, Olympus and *Nature* go to all of the competitors. ■

Graham A. Dunn is at The Randall Centre, Room 3.22D New Hunt's House, King's College London, Guy's Campus, London SE1 1UL, UK.
e-mail: graham.a.dunn@kcl.ac.uk

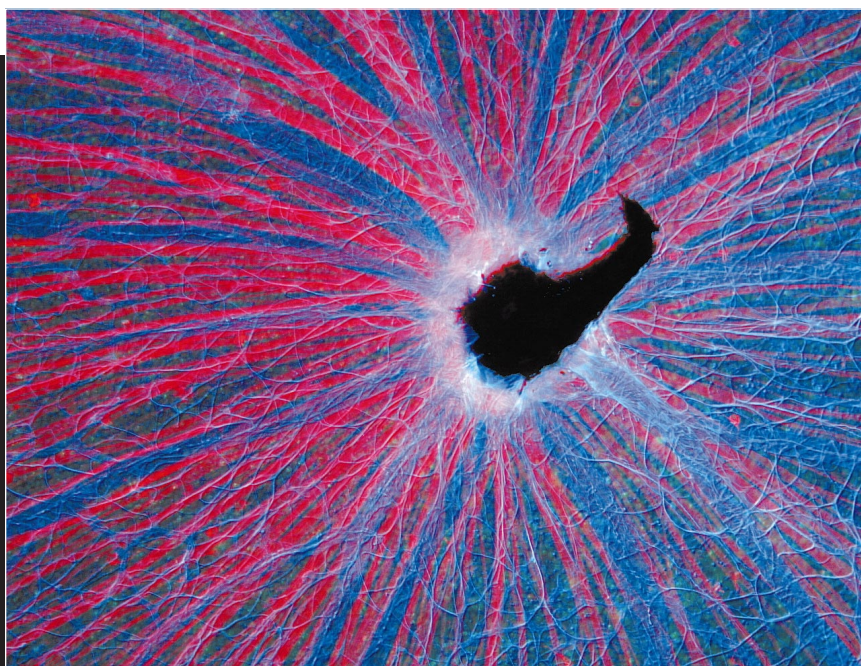


Figure 2 Second prize, Eric Cho. The region around the optic disc of a hamster retina showing axons (red) of retinal ganglion cells converging onto the disc. Blood vessels going into and out of the retina via the optic disc are outlined by the processes of astrocytes (blue) wrapping around them. *Technical details.* Immunofluorescence using anti-neurofilament antibody and anti-GEAP antibody to reveal the ganglion cell axons and astrocytes, respectively. The nuclei of ganglion cells are labelled by applying Diamidino Yellow to the optic nerve. Epifluorescence on Zeiss Axioskop with 10× Neofluar objective and SPOT camera. Magnification, ×82.

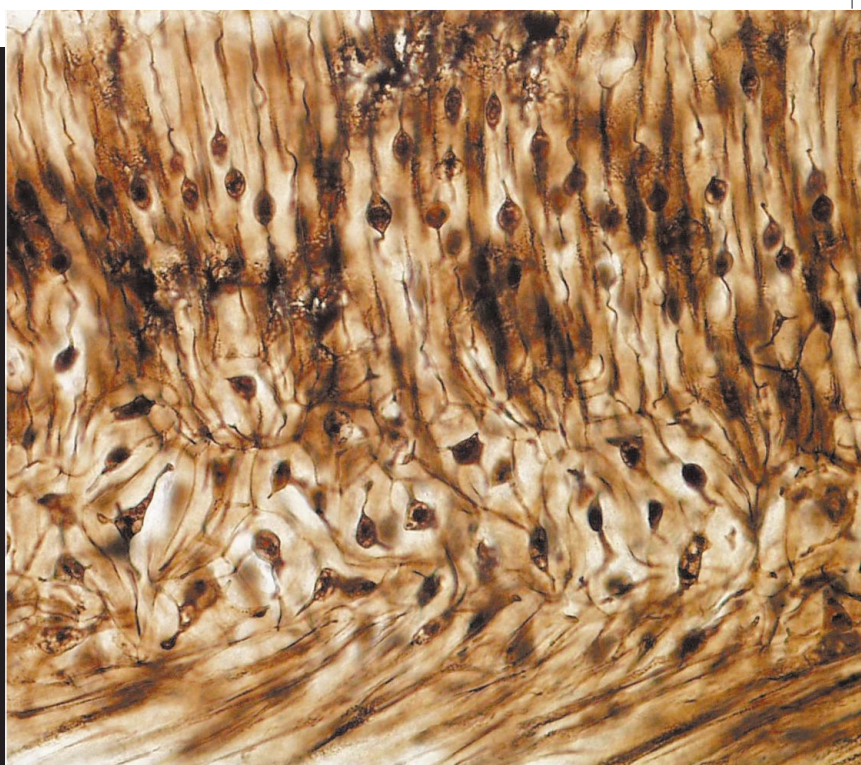


Figure 3 Third prize, Jason Hilton. Part of a fossilized ovule of *Stephanospermum akenoides* Brongn. from the Late Carboniferous of Grand Croix, France. This specimen, which is about 290 million years old, shows exceptional preservation through permineralization by silica. Specimen number 1738 from the Oliver Slide Collection, Natural History Museum, London. *Technical details.* Transmitted light on Olympus BX 40 microscope with Olympus Plan 20× 0.4 na objective and Olympus DP10 digital camera. Magnification, ×257.

Older bull elephants control young males

Orphaned male adolescents go on killing sprees if mature males aren't around.

Musth is a state of heightened sexual and aggressive activity in male elephants^{1,2}. Between 1992 and 1997, young orphaned musth male African elephants (*Loxodonta africana*) that had been introduced to Pilanesberg, South Africa, killed more than 40 white rhinoceros (*Ceratotherium simum*). The killing ceased after six older male elephants were introduced from the relatively normal Kruger Park population. The deviant behaviour of the young Pilanesberg males was rectified by the consequent reduction in musth.

Musth, during which males experience dramatic surges in circulating testosterone^{3,4}, is characterized by a distinct posture, swollen and secreting temporal glands, urine dribbling, and increased sexual and aggressive behaviour². Males gradually enter musth as they become bigger and more experienced and better at winning encounters with other males⁵. In natural populations musth first occurs between 25 and 30 years of age² and its duration increases with age (for age 25–30 years, musth lasts from days to weeks; at 31–35, it lasts for several weeks; at 36–40, for 1–2 months; and after 40 years of age, 2–4 months)².

Young orphan elephants of less than ten years old were reintroduced to Pilanesberg in the 1980s. These were the survivors of Kruger Park culls and they matured in the absence of adult males. Males were breeding successfully by 18 years old and were entering musth from that age. Musth lasted for up to five months, an unusually long time even for males of more than twice that age².

In Amboseli, Kenya, young male elephants were found to be significantly less likely to be in musth if a larger musth male was present⁵. Young males also lose the physical signs of musth minutes or hours

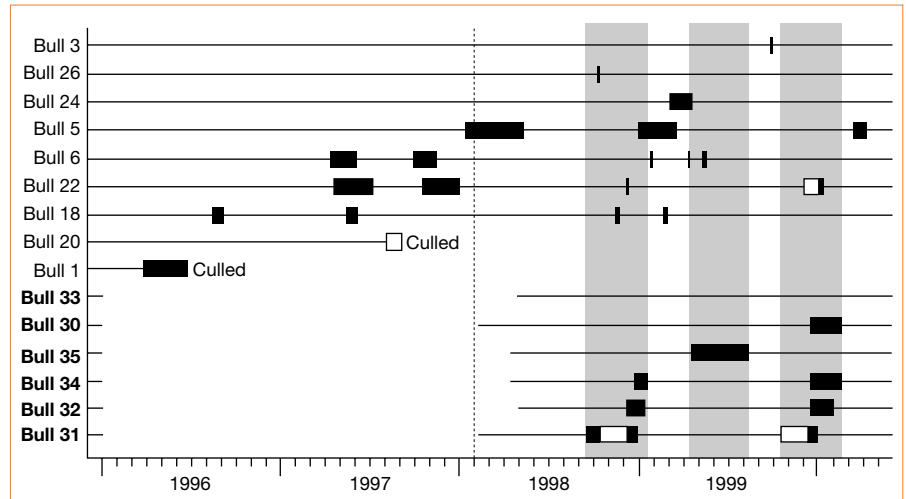


Figure 1 Observed (black bars) and probable (white bars) musth periods of elephants identified by number. Kruger Park males (bulls 30–35, bold type) were introduced to Pilanesberg (25° 15' S, 27° 05' E) between 6 February (vertical line) and 21 March 1998. Shaded areas indicate periods when at least one Kruger male was in musth. Only Pilanesberg males (bull numbers not in bold) that entered musth are shown. The original elephant population was brought from Kruger in 1981 (13 juvenile males, 5 juvenile females), 1983 (13 juvenile males, 11 juvenile females) and 1993 (19 males, 17 females) and from Namibia in 1982 (1 juvenile male, 1 juvenile female). The first calf was born in 1989, showing that males began breeding at about 18 years. Additional details are available from the authors.

after an aggressive interaction with a higher-ranking musth male — it seems that they are forced out of musth by repeated encounters with older males^{2,5}. By inference, larger, older males may delay the onset of musth in younger males.

We tested this idea by introducing six older males to Pilanesberg's population of about 85 elephants, which included 17 young males aged between 15 and 25 and independent of family groups. Musth had been recorded in six of these 17. We expected musth in the original young males to be delayed and, with deviant behaviour being controlled by a reduction in rank, fewer rhinoceros to be killed.

Musth duration in young males declined significantly after the older Kruger Park males were introduced to Pilanesberg (Table 1 and Fig. 1; Wilcoxon paired signed ranks test: $z = -1.826$, 1-tailed $P = 0.034$). Indeed, musth was shorter in all males, and musth periods of males entering musth for the first time were also significantly reduced (Table 1; Mann–Whitney U -test: $z = -1.938$, 1-tailed $P = 0.027$) after the introduction.

The two most extensive periods of musth in the original males occurred when no Kruger Park males were in musth (Fig. 1). Bull 5, the largest original male, was only in musth when Kruger Park males were not; his second musth period was substantially shortened, his third was further reduced, and in 2000 it was delayed by two months (Table 1; Fig. 1).

Despite our small sample sizes, these results provide strong support for the idea that older males suppress musth patterns in younger males⁵. Assessment theory^{6,7} predicts that selection should favour individuals who are able to assess the physical and behavioural traits of rival males and use this knowledge of the costs and benefits of fighting and the probability of winning to adjust their own behaviour.

A strategy of dropping out of musth when confronted with aggression from higher-ranking musth males increases immediate survival, and therefore long-term reproductive fitness. Some experience may be necessary for behavioural control under musth, and its gradual acquisition

Table 1 Effect of introducing older males on musth in young Pilanesberg males

Bull	Musth before introduction		Musth after introduction	
	Dates	Duration (days)	Dates	Duration (days)
1	23 Mar–27 Jun 1996	> 95	Culled before introduction	
18	18 Aug–4 Sep 1996	> 16	17–23 Nov 1998	6
	19 May–1 Jun 1997	> 14	25 Feb 1999	1
6	14 Apr–8 Jun 1997	55	24–26 Jan 1999	3
	29 Sep–13 Nov 1997	46	13 Apr 1999	1
			8–19 May 1999	12–26
22	Apr–Jul 1997	~90	3–4 Dec 1998	2
	12 Oct 1997–6 Jan 1998	> 86	20 Nov 1999–12 Jan 2000	53–77
20	Aug 1997	Unknown	Culled before introduction	
5	4 Jan–18 May 1998	134	6 Jan–21 Mar 1999	76
			12 Mar–5 Apr 2000	24
26	None		3–9 Oct 1998	6
24	None		2 Mar–24 Apr 1999	53
3	None		30 Sep 1999	1

There may have been undocumented periods of musth before April 1996. 'Greater than' sign indicates minimum periods when bulls were not seen immediately before or after the specified dates. Ranges indicate the likely interval between the first or last date the elephant was seen in musth, and the date when it was seen out of musth before or after those musth dates. Numbers in bold were used in paired analysis.

may acclimatize males physiologically and psychologically to the extremely high levels of circulating testosterone^{3,4}.

This translocation ended the killing of rhinoceros by elephants. Introducing higher-ranking males may help in other reserves containing Kruger orphans where managers face similar problems. In elephant populations that have experienced massive poaching, the hunters' selection of larger males may cause similar problems once rhinoceros populations increase, and these too may be helped by importing older male elephants.

Rob Slotow*, Gus van Dyk†, Joyce Poole‡, Bruce Page*, Andre Klocke*§

*School of Life and Environmental Science, George Campbell Building, University of Natal, Durban 4041, South Africa

e-mail: slotow@biology.und.ac.za

†North West Parks and Tourism Board, Pilanesberg National Park, PO Box 1305 Mogwase, South Africa

‡PO Box 24747, Nairobi, Kenya

§Deceased

1. Poole, J. H. & Moss, C. J. *Nature* **292**, 830–831 (1981).
2. Poole, J. H. *Behaviour* **102**, 283–316 (1987).
3. Poole, J. H., Kasman, L. H., Ramsay, E. C. & Lasley, B. L. *J. Reprod. Fert.* **70**, 255–260 (1984).
4. Rasmussen, L. E. L., Hall-Martin, A. J. & Hess, D. L. *J. Mammal.* **77**, 422–439 (1996).
5. Poole, J. H. *Anim. Behav.* **37**, 140–152 (1989).
6. Parker, G. A. *J. Theor. Biol.* **47**, 223–243 (1974).
7. Maynard-Smith, J. & Parker, G. A. *Anim. Behav.* **24**, 159–175 (1976).

Materials science

C₆₆ fullerene encaging a scandium dimer

The geometry of carbon cages (fullerenes) is governed by the isolated-pentagon rule (IPR)^{1,2}, which states that the most stable fullerenes are those in which all pentagons are surrounded by five hexagons. Although this rule has been verified experimentally^{3–5}, it is impossible for fullerenes in the range C₆₀ to C₇₀ to obey it. Here we describe the production and characterization of an IPR-violating metallofullerene, Sc₂@C₆₆, a C₆₆ fullerene encaging a scandium dimer. Our results indicate that encapsulation of the metal dimer significantly stabilizes this otherwise extremely unstable C₆₆ fullerene.

We generated soot containing Sc₂@C₆₆ and other scandium metallofullerenes^{3,4} in direct-current arc discharge of scandium/graphite composite rods under a flow of helium. Sc₂@C₆₆ metallofullerene was isolated using multistage high-performance liquid chromatography⁶. Starting with 800 g arc-generated soot, we were able to isolate about 2.0 mg Sc₂@C₆₆. The purity (99.8%) of this material was confirmed by laser-desorption time-of-flight mass spectrometry.

There are several ways in which the IPR can be violated, the most straightforward being to generate the so-called 'fused-pentagon' in which pentagons are adjacent to one another. For 66-atom carbon cages with hexagonal and pentagonal faces, there are 4,478 possible (non-IPR) structural isomers with 2 × D_{3h}, 1 × C_{3v}, 18 × C_{2v}, 112 × C_s, 211 × C₂ and 4,134 × C_i symmetry⁷. Considering the observed 19-lines (5 × 2; 14 × 4) in the high-resolution ¹³C NMR spectrum of Sc₂@C₆₆ (Fig. 1a), only eight structural isomers of C₆₆ with C_{2v} symmetry can satisfy this ¹³C NMR pattern.

To determine the geometrical structure of the metallofullerene unambiguously, we made synchrotron X-ray powder diffraction

measurements on Sc₂@C₆₆ at Spring-8 BL02B2. The MEM (maximum-entropy method)^{5,8} three-dimensional electron-density distribution of Sc₂@C₆₆ obtained by the MEM/Rietveld procedure^{5,8} using synchrotron X-ray diffraction is shown in Fig. 1b, together with an optimized geometry of Sc₂@C₆₆ based on non-local density-functional B3LYP/Basis set [Sc(LanL2DZ); C(3-21G)] calculations (Fig. 1c).

The Sc₂@C₆₆ X-ray crystal structure is of space group *Pmn*2₁(no. 31), where *a* is 10.552(2) Å, *b* is 14.198(2) Å, *c* is 10.553(1) Å, and the reliability factors of the final Rietveld fitting were *R*_{wp} = 2.4% and *R*_f = 13.1%. The final charge densities obtained by the maximum-entropy method, whose reliability factor is *R*_F = 5.4%, reveal a pair of two-fold fused pentagons on a C₆₆-C_{2v} cage that encapsulates a Sc₂ dimer; the most stable Sc₂@C₆₆ structure has the

least number and degree of fused pentagons out of the 4,478 possible isomers (see Supplementary Information).

The ¹³C NMR spectrum of Sc₂@C₆₆ shows five highly deshielding lines above 151 p.p.m. up to 167.9 p.p.m. These unusual ¹³C NMR lines stem from those carbon atoms on the fused-pentagon area. In particular, the two lowest lines at half intensity (numbers 16 and 17 in Fig. 1a) originate from the four carbon atoms that connect two pentagonal rings.

The Sc₂@C₆₆ structure (Fig. 1c) contains two pairs of two-fold fused pentagons with two closely situated scandium atoms. The observed Sc–Sc distance is 2.87(9) Å, indicative of the formation of a Sc₂ dimer inside the C₆₆ cage. Endohedral metallofullerenes are stabilized by intrafullerene electron transfer^{3–5,9,10}; there are 40.0(2) *e* electrons in the area corresponding to the Sc₂ dimer calculated from the MEM charge density, a value that is very close to (Sc₂)²⁺ with 40 *e*. The *ab initio* calculation also indicates that the Sc₂ dimer donates two electrons to the C₆₆ cage, resulting in a formal electronic state of (Sc₂)²⁺@C₆₆^{2–}. It is this charge-transfer interaction between the Sc₂ dimer and the fused pentagons that significantly decreases the strain energies caused by the pair of fused pentagons and thus stabilizes the fullerene cage. We conclude that the isolated-pentagon rule is not necessarily a test for stable geometry in endohedral metallofullerenes⁹.

Chun-Ru Wang*, Tsutomu Kai*, Tetsuo Tomiyama*, Takuya Yoshida†, Yuji Kobayashi†, Eiji Nishibori‡, Masaki Takata‡, Makoto Sakata‡, Hisanori Shinohara*

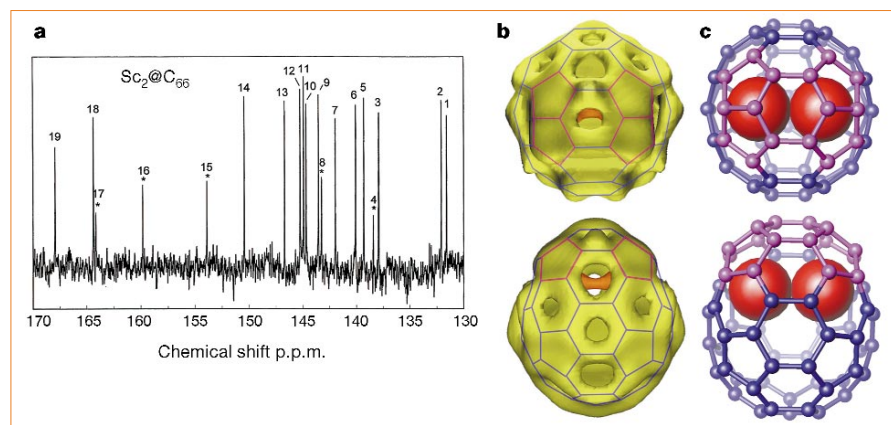


Figure 1 ¹³C NMR and synchrotron X-ray powder diffraction structural data on Sc₂@C₆₆. **a**, ¹³C NMR spectrum of Sc₂@C₆₆ in CS₂ solution (2.5 mg Cr(acac)₃, where 'acac' represents acetylacetonate, relaxant and C₆D₆ lock) after 160,000 scans at room temperature using a Varian Inova 600 spectrometer at 600 MHz. The five lines marked with an asterisk are half the intensity of the other 14 full lines (131.6, 132.1, 137.9, 138.4*, 139.3, 140.1, 142.0, 143.3*, 143.5, 144.7, 144.9, 145.2, 146.7, 150.4, 153.9*, 159.8*, 164.1*, 164.4 and 167.9 p.p.m.). **b**, X-ray structure of the IPR-violating Sc₂@C₆₆ fullerene, showing a top view along the C₂ axis and a side view. The equi-contour (1.4 e Å⁻³) surface of the final maximum-entropy method electron charge density is shown; the Sc₂ dimer is in red and the two pairs of fused pentagons are evident. The X-ray powder pattern had good counting statistics and was measured in an X-ray powder experiment using synchrotron radiation and imaging-plate detectors exposed for 2 h (incident X-ray wavelength, 0.75 Å). The X-ray powder pattern of Sc₂@C₆₆ was obtained with a 0.02° step up to 20.3° in 2θ, which corresponds to 2.0 Å resolution in *d* spacing. **c**, Calculated Sc₂@C₆₆ structures (see text).

*Department of Chemistry, Nagoya University,
Nagoya 464-8602, Japan

e-mail: nori@nano.chem.nagoya-u.ac.jp

†Department of Pharmacology, Osaka University,
Osaka 565-0871, Japan

‡Department of Applied Physics, Nagoya
University, Nagoya 464-8603, Japan

1. Kroto, H. *Nature* **329**, 529–531 (1987).
2. Schmalz, T. G., Seitz, W. A., Klein, D. J. & Hite, G. E. *J. Am. Chem. Soc.* **110**, 1113–1127 (1988).
3. Shinohara, H. *et al.* *Nature* **357**, 52–54 (1992).
4. Yannoni, C. S. *et al.* *Science* **256**, 1191–1192 (1992).
5. Takata, M. *et al.* *Nature* **377**, 46–49 (1995).
6. Shinohara, H. *et al.* *J. Phys. Chem.* **97**, 4259–4261 (1993).
7. Fowler, P. W. & Manolopoulos, D. E. *An Atlas of Fullerenes* 27–42 (Clarendon, Oxford, 1995).
8. Takata, M. *et al.* *Phys. Rev. Lett.* **83**, 2214–2217 (1999).
9. Kobayashi, K., Nagase, S., Yoshida, M. & Osawa, E. *J. Am. Chem. Soc.* **119**, 12693–12694 (1997).
10. Dorn, H. C. *et al.* in *Fullerenes: Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials* (eds Kadish, K. M. & Ruoff, R. S.) 990–1002 (Electrochemical Society, Pennington, 1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Materials science

A stable non-classical metallofullerene family

In the evolving field of fullerenes, nanotubes and endohedral metallofullerenes, the isolated-pentagon rule (IPR)¹ is sacrosanct — exceptions have been predicted^{2,3}, but no bare carbon cages with adjacent pentagons have been characterized. Small organic molecules with metal-stabilized fused five-membered rings (pentalenes) have been created⁴, however, and here we describe a family of non-classical endohedral metallofullerenes with the general structure $A_xSc_{3-x}N@C_{68}$ (where $x=0-2$, A is a rare-earth metal, Sc is scandium and N is nitrogen) that has a non-IPR cage of only 68 carbon atoms containing annelated five-membered rings. This internal ring network is metal-stabilized and is accessible for external organic reaction chemistry.

We formed the scandium encapsulate $Sc_3N@C_{68}$ in a Krätschmer–Huffman⁵ generator using the trimetallic nitride template (TNT, in the presence of nitrogen) process, purifying the encapsulate by chromatography⁶. We obtained $Sc_3N@C_{68}$ (Fig. 1a) with soluble extract yields of 1–2%; this is better than the most abundant non-TNT scandium encapsulate, $Sc_2@C_{84}$ and is only exceeded by the $Sc_3N@C_{80}$ family.

It is easy to prepare other non-IPR family members, $A_xSc_{3-x}N@C_{68}$ ($x=0-2$) from graphite rods containing a mixture of scandium and rare-earth metal (A) oxide⁶, as our chromatographic isolation of a purified sample of $ErSc_2N@C_{68}$ at levels similar to $Sc_3N@C_{68}$ shows (Fig. 1b). Yields of this mixed scandium/rare-earth family for $ASc_2N@C_{68}$ and $A_2ScN@C_{68}$, where A can

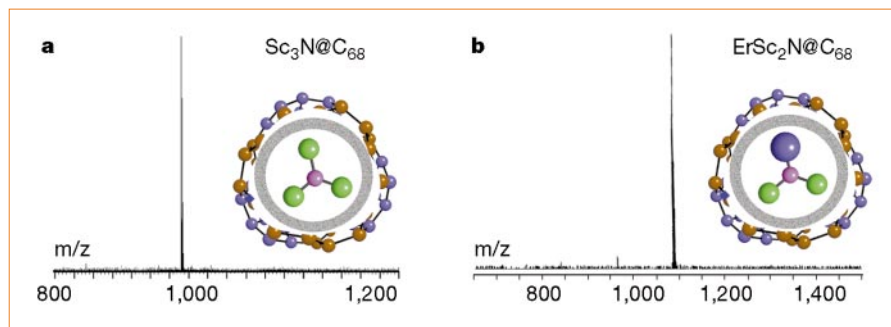


Figure 1 Mass-spectral data for purified $Sc_3N@C_{68}$ and $ErSc_2N@C_{68}$. **a**, The Ni-DCl mass spectrum for $Sc_3N@C_{68}$, with $m+1$, $m+2$ and $m+3$ isotope intensities have predicted (observed) values of 76.70 (76.8 ± 2.2), 28.98 (28.0 ± 1.2) and 7.19 (7.6 ± 0.5), respectively. **b**, The Ni-DCl mass spectrum for purified $ErSc_2N@C_{68}$, with m , $m+1$, $m+2$ and $m+3$, have predicted (observed) isotope intensities of 63.3 (62.2 ± 4.8), 90.3 (84.2 ± 8.5), 100 (100) and 54.4 (53.0 ± 4.2), respectively. The 150-MHz ^{13}C -NMR spectrum for $Sc_3N@C_{68}$ consists of signals at 158.49, 150.40, 149.51, 147.41, 145.52, 143.55, 142.92, 137.77, 137.62, 137.19, 137.12 and 136.87 (1/3 intensity) p.p.m. The ^{45}Sc and ^{13}C NMR, UV-visible, and chromatographic data for $Sc_3N@C_{68}$ are given in Supplementary Information.

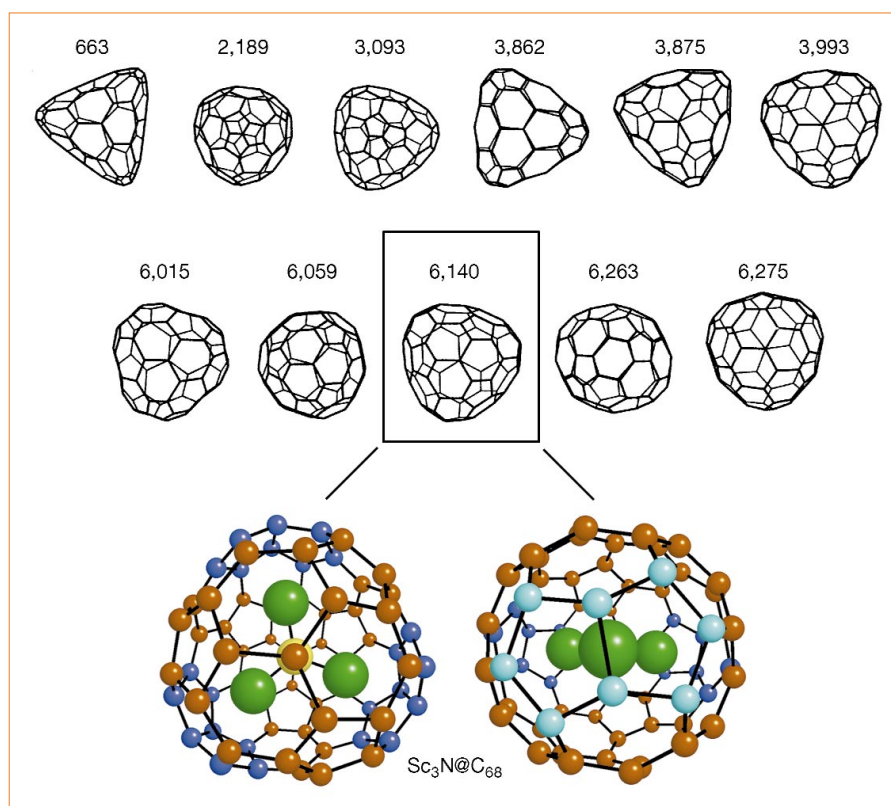


Figure 2 Three-fold symmetry isomers of C_{68} and proposed $Sc_3N@C_{68}$ structure (isomer 6,140). There are 11 non-IPR isomers generated from the spiral algorithm⁷ with three-fold symmetry axis consistent with 12 ^{13}C NMR lines (11 of equal intensity and 1 of one-third intensity) and a single symmetric ^{45}Sc NMR signal. Isomer 663, D_3 , $N_p=15$; isomers 2,189, 3,093, 3,862, 3,875 and 3,993, D_3 , $N_p=9$; isomers 6,015 and 6,059, D_3 , $N_p=6$; isomer 6,263, S_6 , $N_p=6$; isomers 6,140 and 6,275, D_3 , $N_p=3$. Isomer 6,140 has spiral $5^2 6^6 (56)^2 6 (56)^2 6^4 5 6^4 (56)^5$. Two orthogonal views are shown for the proposed $Sc_3N@C_{68}$ structure generated from isomer 6,140 geometry (DFTB computation level) with encapsulation of the Sc_3N cluster. Views are shown from the C_3 axis and the C_2 axis.

be Tm, Er, Gd, Ho or La, were more modest (see Supplementary Information). The absence of measurable quantities of endohedral clusters of three larger rare-earth atom clusters ($A_3N@C_{68}$) is consistent with the much larger radii of rare-earth atoms compared with Sc ions in a space-confining cage of only 68 carbon atoms.

For fullerenes smaller than C_{70} , only C_{60} , with its classic icosahedral cage, has an isomer allowed by the IPR¹. For a C_{68} cage, the spiral algorithm⁷ finds 6,332 distinct

fullerenes containing pentagon and hexagon rings, but only 11 of these structures are consistent with the single symmetric line and 12 singlet peaks (11 of equal intensity and 1 of one-third intensity) seen in the ^{45}Sc and ^{13}C NMR spectra, respectively, for $Sc_3N@C_{68}$ (see Supplementary Information). All candidate isomers have either D_3 or S_6 symmetry (Fig. 2). From the NMR data and assuming three-fold symmetry, other defect rings, such as heptagons, in alternative cages would have to occur in sets

of three, with correspondingly larger energy penalties^{2,8}. Each pair of fused pentagons in a neutral fullerene cage carries an energy penalty of 70–90 kJ mol⁻¹ with respect to the structure of C₆₀ (ref. 8). Two candidate structures (isomers 6,140 and 6,275) have the minimum three fused pentagon pairs; all others in the set have from 6 to 15 pentagon fusions.

We confirmed the qualitative preference for isomers with low numbers of fused pentagons using model calculations that treat the cage as an empty fullerene capable of accepting electrons from a central reservoir. At the density-functional tight binding level of computation⁹, with full geometry optimization of a closed-shell electron configuration, the empty cage 6,140 is stabilized by 120 kJ mol⁻¹ with respect to its nearest rival, with a 12-line NMR spectrum and the minimal three pentagon adjacencies (isomer 6,275). As between two and six excess electrons are added to the cage, to simulate the range of likely charge transfer from the encapsulated cluster, isomer 6,140 becomes increasingly favoured over all other empty cages in the set. In view of this consistent preference, we propose a structure for Sc₃N@C₆₈ (Fig. 2) consisting of the encapsulated Sc₃N cluster in the C₆₈ (D₃) three-fold symmetric isomer 6,140 cage. The encapsulated Sc₃N cluster is shown with the Sc atoms on C₂ axes, but from the ⁴⁵Sc NMR it is also possible that these represent time-averaged orientations.

S. Stevenson*, P. W. Fowler†, T. Heine‡, J. C. Duchamp§, G. Rice*, T. Glass*, K. Harich*, E. Hajdull, R. Biblell, H. C. Dorn*

*Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061, USA

e-mail: hdorn@chemserver.chem.vt.edu

†School of Chemistry, University of Exeter, Stocker Road, Exeter EX4 4QD, UK

‡Dipartimento di Chimica 'G. Ciamician', Università di Bologna, via Selmi 2, Bologna I-40126, Italy

§Department of Chemistry, Emory and Henry College, Emory, Virginia 24327-0943, USA

||Searle, 4901 Searle Parkway, Skokie, Illinois 60077, USA

1. Kroto, H. W. *Nature* **329**, 529–531 (1987).
2. Kobayashi, K., Nagase, S., Yoshida, M. & Osawa, E. *J. Am. Chem. Soc.* **119**, 12693–12694 (1997).
3. Dorn, H. C. et al. in *Fullerenes: Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials* (eds Kadish, K. M. & Ruoff, R. S.) 990–1002 (Electrochemical Society, Pennington, 1998).
4. Butenschön, H. *Angew. Int. Edn Engl.* **36**, 1695–1697 (1997).
5. Krätschmer, W., Fostiropoulos, K. & Huffman, D. R. *Chem. Phys. Lett.* **170**, 167–170 (1990).
6. Stevenson, S. et al. *Nature* **401**, 55–57 (1999).
7. Fowler, P. W. & Manolopoulos, D. E. *An Atlas of Fullerenes* (Oxford Univ. Press, Oxford, 1995).
8. Albertazzi, E. et al. *Phys. Chem. Chem. Phys.* **1**, 2913–2919 (1999).
9. Seifert, G., Porezag, D. & Frauenheim, T. *Int. J. Quantum Chem.* **58**, 185–192 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Palaeontology

The 'feathers' of *Longisquama*

The elongated dorsal appendages of the reptile *Longisquama insignis*, from the Triassic of Kyrgyzstan¹, have recently been reinterpreted as the first record of feathers in a non-avian tetrapod² — long predating the feathers of the oldest known bird, *Archaeopteryx*. Here we present evidence that the dorsal scales of *Longisquama* are not feathers, and that they are in fact strikingly different from avian feathers. We conclude that *Archaeopteryx* remains the oldest known feathered tetrapod.

Longisquama is a small diapsid reptile with an uncertain phylogenetic position. It is known from an incomplete skeleton with integumentary appendages and isolated appendages. Appendage PIN (for Palaeontological Institute of the Russian Academy of Sciences) 2584/7, preserved as part and counterpart, retains an infilling of fine-grained sediment and high-fidelity impressions of the external left and right surfaces of the appendage (Fig. 1). This infilling, preserved either on one side of the specimen or on the counterpart, shows that the tubular configuration described for the proximal portion extends along the entire length of the appendage, although the distal portion is expanded anteroposteriorly and flattened transversely. This indicates that in life the two external surfaces were separated from each other by an intervening space (now sediment-filled).

There are no feather-like features on the distal portion of the appendage. Here, two corrugated membrane-like surfaces touch along their leading and trailing edges to form wide, smooth bands. The two membranes were apparently supported by a median vein-like structure extending the length of the appendage. This has been proposed as the homologue of the rachis of avian feathers². On either side of this 'vein', the external surfaces of the appendage are corrugated. This corrugation varies along the appendage: proximally, individual rugae are relatively large and widely spaced, but in the distal portion they are smaller and densely packed. The densely arranged distal corrugations have been compared to the pinnae of avian feathers², but the fossils indicate that these are formed on a membrane-like structure on either side of the 'vein'.

The fossils were split into part and counterpart during collecting, and most of the appendages are now preserved as impressions of their left and right sides, without the intervening sediment core. The surfaces of both the part and counterpart impressions of individual appendages are concave, an indication that these structures are three-dimensional. In contrast, the parts

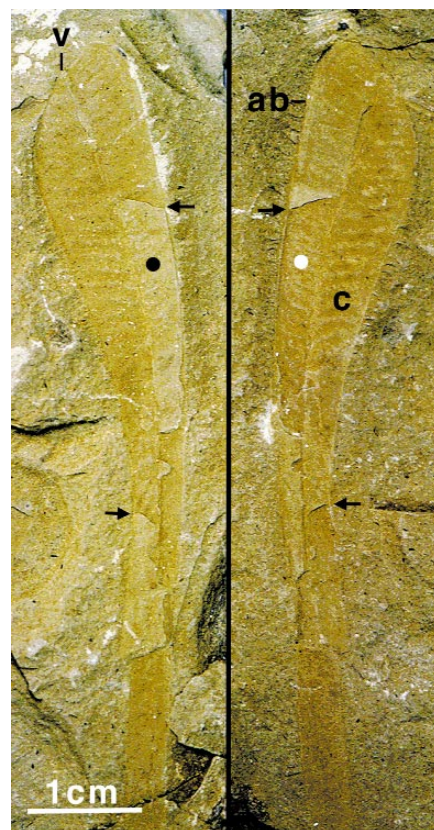


Figure 1 Part and counterpart of an elongated dorsal scale of *Longisquama insignis* (PIN 2584/7). Where the sedimentary infilling (black circles) is not preserved, sharp impressions of the corrugated external surface of the structure are visible (white circles). Arrows point to corresponding patches of sedimentary infilling on part and counterpart. ab, anterior smooth band; c, corrugations; v, median 'vein'.

and counterparts of feather impressions in *Archaeopteryx* are concave and convex, respectively.

We believe that the dorsal appendages of *Longisquama* are highly modified scales, as suggested previously^{1,3}, rather than feathers. Examination of the holotype of *L. insignis* (PIN 2584/4) suggests that they were anchored in the skin or epaxial muscles.

Robert R. Reisz*, Hans-Dieter Sues†

*Department of Biology, University of Toronto in Mississauga, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada

e-mail: rreisz@credit.erin.utoronto.ca

†Department of Palaeobiology, Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada

e-mail: hdsues@rom.on.ca

1. Sharov, A. G. *Paleontol. Zhur.* **1970**, 127–130 (1970).
2. Jones, T. et al. *Science* **288**, 2202–2205 (2000).
3. Feduccia, A. *The Origin and Evolution of Birds* (Yale Univ. Press, New Haven, 1996).

Correction

Detection of preinvasive cancer cells

V. Backman et al.

Nature **406**, 35–36 (2000)

The name of the tenth author of this communication is J. A. McGilligan (not T. McGillican as published).

In search of the tumour-suppressor functions of BRCA1 and BRCA2

Ralph Scully & David M. Livingston

The Dana-Farber Cancer Institute and the Harvard Medical School, Boston, Massachusetts 02115, USA

Hereditary breast and ovarian cancer syndromes can be caused by loss-of-function germline mutations in one of two tumour-suppressor genes, *BRCA1* and *BRCA2* (ref. 1). Each gene product interacts with recombination/DNA repair proteins in pathways that participate in preserving intact chromosome structure. However, it is unclear to what extent such functions specifically suppress breast and ovarian cancer. Here we analyse what is known of *BRCA* gene function and highlight some unanswered questions in the field.

The amino acid sequences of BRCA1 and BRCA2 originally provided few clues to their specific functions (Fig. 1a). BRCA1 was found to localize to sub-nuclear foci during the S and G2 phases of the cell cycle. These foci also contain Rad51, a protein central to homologous recombination². Endogenous BRCA1-Rad51 complexes were detected in cell extracts. BRCA2 also interacts and co-localizes with Rad51 and BRCA1 (refs 3 and 4). The three genes are co-expressed in development¹.

Human Rad51, like its homologues in other eukaryotes (RAD51) and prokaryotes (RecA), operates in the repair of double-strand DNA breaks (DSB) by binding single-stranded (ss) DNA to form a nucleoprotein filament that can invade a homologous duplex DNA molecule. *Saccharomyces cerevisiae rad51* mutants are defective in double-strand break repair (DSBR) and meiosis, two processes dependent upon homologous recombination.

Rad51 also exhibits characteristic localization patterns during meiosis. The synaptonemal complex (SC) is a linear protein/DNA-bearing structure that assembles at the onset of meiotic prophase I, a tetraploid stage during which homologous chromosomes become paired before meiotic recombination. The SC forms by the apposition ('synapsis') of two homologous axial elements, each containing two identical sister chromatids and associated proteins. Rad51 binds axial elements. Strikingly, BRCA1 and BRCA2 act similarly^{2,4} (Fig. 1b), suggesting that BRCA proteins function in meiotic recombination, and hinting at a role in sister chromatid interactions.

Relocation of BRCA1 to sites of DNA synthesis

After hydroxyurea (HU)-mediated replication arrest or exposure of S-phase cells to ultraviolet (UV) light, BRCA1 rapidly relocates from its 'native' nuclear foci to sites of DNA synthesis and becomes hyperphosphorylated⁵. This suggests that BRCA1 is involved in the repair of abnormal DNA structures generated at sites of DNA synthesis after DNA damage. Among the proteins that interact and co-localize with BRCA1 in nuclear foci are BRCA2 and BARD1. Like BRCA1, BARD1 contains an amino-terminal RING domain and carboxy-terminal BRCT motifs⁶. BARD1, BRCA2 and Rad51 all accompany BRCA1 in its re-localization after DNA damage^{4,5}. This implies that these multiprotein complexes are dedicated, at least in part, to repair of replication-associated DNA damage.

In yeast, cell cycle control responses to HU or to certain DNA damaging agents delivered in the S phase are controlled by S-phase checkpoint gene products, which include *S. cerevisiae* MEC1 and *Schizosaccharomyces pombe* rad3, two large nuclear kinases with C-terminal PI3 kinase-like domains. Atr is a mammalian homologue of these proteins and, like BRCA proteins, it also localizes to axial elements during meiotic prophase I⁷. Recently, Atr was found to colocalize with BRCA1 in somatic cells, both before and after

replication arrest⁴¹. Atr, in part, controls BRCA1 phosphorylation following HU treatment⁴¹. BRCA1 is also phosphorylated following gamma irradiation⁵, under the control of a related kinase, Atm⁸.

'Death by checkpoint'

Mouse gene targeting experiments provided insights into the relationship between BRCA mutation and cancer (reviewed in refs 1 and 9). Nullizygous *BRCA1* or *BRCA2* embryos die around the time of gastrulation. Unexpectedly, these embryos reveal a proliferative defect and induction of the *p53*-dependent cell cycle inhibitor, *p21* (ref. 10). It was argued that, if BRCA1 regulates DNA repair, its inactivation might lead to spontaneous abnormalities in DNA structure. If so, induction of *p21* in *BRCA1*^{-/-} embryos might reflect the activation of a DNA damage-dependent checkpoint². The resulting cell cycle delay could have catastrophic effects on a gastrulating embryo, leading to 'death by checkpoint'. Consistent with this hypothesis, *p53* or *p21* nullizygosity delays the death of *BRCA1*^{-/-} or *BRCA2*^{-/-} embryos (reviewed in refs 1 and 9).

BRCA1 dysfunction might have similar effects in adult cells; it may precipitate spontaneous abnormalities of DNA structure and, thereby, provoke 'death by checkpoint'. If, however, loss of *BRCA1* function were to occur in a pre-malignant cell that had already suffered inactivation of key checkpoints, the aberrant DNA structures resulting from *BRCA1* inactivation might be tolerated without cell cycle arrest² (Fig. 2). This might promote neoplastic development. *BRCA1* or *BRCA2* inactivation, if it were to lead to cancer, would then not be the first 'hit' during tumorigenesis. Alternatively, *BRCA* gene inactivation in an otherwise wild-type cell would have to be closely followed by the inactivation of key checkpoints.

Support for the notion that *BRCA* genes suppress genome instability comes from several sources. Embryonic tissue lacking *wt* (wild-type) *BRCA1* or *BRCA2* reveals ionizing radiation (IR) hypersensitivity, consistent with a defect in DSBR^{3,9}. *BRCA1* or *BRCA2* homozygous mutant mouse embryo fibroblasts (MEFs) undergo spontaneous chromosome breakage accompanied by checkpoint-mediated growth arrest^{9,11}. Moreover, severe aneuploidy and centrosome amplification are observed in these cells^{1,9}. The full identity of the checkpoints responsible for cell cycle arrest of *BRCA1* or *BRCA2* mutated cells is unclear. Among the proposed candidates is the spindle pole checkpoint^{1,9}. *Rad51*^{-/-} mouse embryos also suffer early death that is partially rescued by *p53* germline inactivation, and cultured cell lines deprived of Rad51 undergo spontaneous chromosome breakage.

Interestingly, *BRCA* gene and checkpoint-mediated mechanisms may also suppress genome instability in breast tissue. Recently, breast-specific murine *BRCA1* gene inactivation was achieved⁹. These mice developed late-onset breast cancer with frequent, tumour-associated *p53* mutations. On a *p53*^{+/-} background,

earlier-onset and higher-frequency disease was detected. *p53* hemizygosity was also found to promote breast tumour development in *BRCA1*^{+/-} mice exposed to ionizing radiation¹⁹. Thus, *p53*-mediated checkpoint functions may contribute to proliferation control of *BRCA1*-deficient breast ductal cells. Loss of these *p53* functions could increase tolerance of genome instability and promote tumorigenesis.

BRCA-mediated recombination and tumour suppression

Human genetics has defined numerous mutant *BRCA1* and *BRCA2* alleles defective in tumour suppression. This has provided a tool for testing the relationship between *BRCA1* tumour suppressor function and its role in genome integrity maintenance. Recently, a breast cancer cell line, HCC1937, that lacks *wtBRCA1* was found to be hypersensitive to IR and to exhibit delayed kinetics of DSBR. Stable expression of *wtBRCA1* in these cells reversed IR sensitivity and restored efficient DSBR¹³. In contrast, several clinically important mutant alleles, each bearing a missense mutation in one of three distinct *BRCA1* protein domains—the N-terminal RING domain, the C-terminal BRCT motifs, or a centrally located region—failed to rescue the phenotype. The DSBR function of *BRCA1*, therefore, arises from the participation of diverse structural elements of the protein. As each of these domains also serves as an interaction centre for associated proteins, a scaffolding role for *BRCA1* in DSBR seems possible. Moreover, given the inactivation of this function by a diverse set of disease-producing, missense mutations, DSBR may be a key *BRCA1* tumour-suppressor function.

At least two distinct processes, homologous recombination (HR) and non-homologous end joining (NHEJ), contribute to DSBR in mammalian cells. The interaction of *BRCA1* with

the NBS1–MRE11–Rad50 complex (reviewed in ref. 1) potentially implicates it in either of these processes. Current data suggest a specific contribution of *BRCA* genes to HR. For example, *BRCA1* mutant mice reveal arrested spermatogenesis in meiotic prophase I, the stage at which *BRCA1* is normally localized to the axial elements of developing synaptonemal complexes (reviewed in refs 1 and 9). A unique, viable *BRCA1* homozygous mutant murine embryonic stem (ES) cell clone revealed reduced rates of HR or single-strand annealing in response to a site-specific DSB¹⁴. Gene targeting, an HR-dependent process, was enhanced by expression of *wtBRCA1* in the same *BRCA1*-mutated ES cell clone¹⁵. By contrast, where measured, NHEJ appeared normal in *BRCA1* or *BRCA2* mutant cells^{11,14}.

DNA polymerase stalling and S-phase checkpoint activation

If the *BRCA* genes act as tumour suppressors by supporting HR, which HR-dependent process is the key disease-relevant target? One candidate is sister chromatid recombination (SCR), a potentially error-free process which operates, in part, in the repair of recombinogenic DNA lesions generated during S phase¹⁶. SCR may be provoked by attempted replication across lesions that induce DNA polymerase stalling, such as those caused by UV irradiation or DNA adduction¹⁷. Indeed, work in prokaryotes indicates that replication arrest is a common event during DNA synthesis and that recombination functions are required for replication restart, genome integrity maintenance, and cell viability (reviewed in ref. 18).

Hydroxyurea exposure may mimic such DNA polymerase-stalling events by causing genome-wide replication arrest. HU or UV treatment of S phase cells may each give rise to persistent ssDNA tracts—‘daughter strand gaps’ (DSG)—in close proximity to a replication fork¹⁹ (Fig. 3). We propose that such DSGs contribute to the rapid relocalization of *BRCA1* and associated proteins to sites of replication following HU or UV treatment⁵. DSGs, or their derivatives, might activate S phase checkpoint signalling and also serve as substrates of recombinational responses, such as SCR (Fig. 3). *Atr* now appears to be a regulator of these events⁴¹. Interestingly, *ATR*^{-/-} embryos suffer early embryonic lethality with spontaneous chromosome breakage—a phenotype reminiscent of *BRCA*^{-/-} and *Rad51*^{-/-} embryos²⁰.

Whereas S-phase ‘checkpoint’ kinases signal to *BRCA1*, little is known of how these phosphorylations alter *BRCA1* function. We do know that, unlike *Atr*, intact *BRCA* proteins are not required for

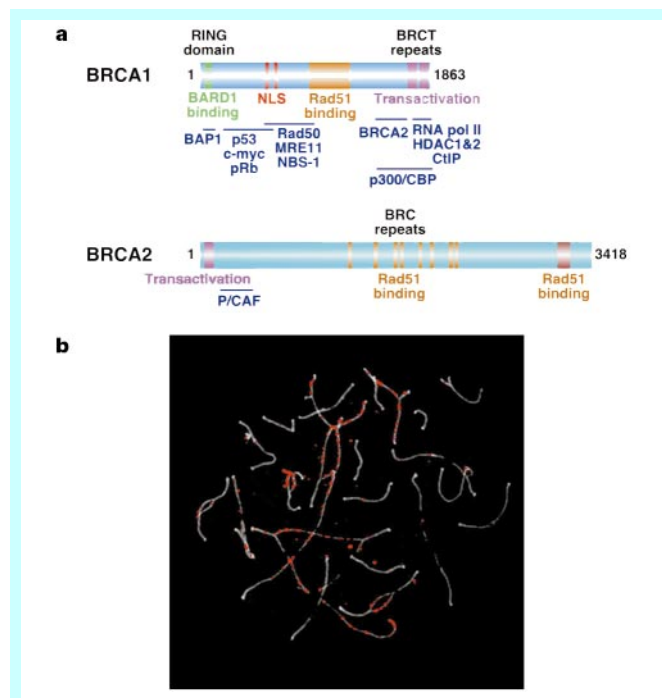


Figure 1 Structure and localization of *BRCA* proteins. **a**, Structure of the *BRCA1* and *BRCA2* polypeptides. *BRCA1* shows RING domain, BRCT motifs, implicated in DNA damage response pathways. *BRCA2* shows BRC repeats, involved in the Rad51 interaction. Some associated proteins are denoted in blue. **b**, Localization of *BRCA1* during meiotic prophase I. Depicted is a single primary, human spermatocyte nucleus, stained with an antibody to the synaptonemal complex protein, SCP3 (white), and with *BRCA1* antibody (red). *BRCA1* localizes to the unsynapsed regions (axial elements) of developing synaptonemal complexes. Copyright held by Cell Press, reproduced with permission.

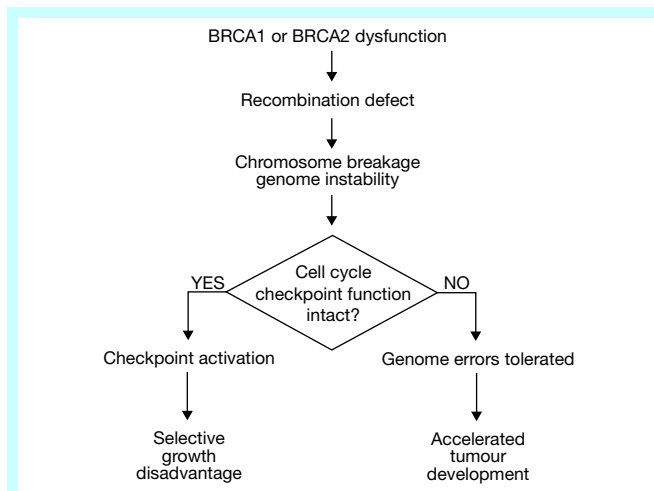


Figure 2 Checkpoint inactivation and *BRCA* gene-mediated tumorigenesis. Inactivation of DNA damage-responsive cell cycle checkpoints may contribute to the cellular response to *BRCA* gene dysfunction. Loss of *BRCA* function and associated chromosome breakage in a cell trigger checkpoints that deselect it. If such a checkpoint has already been disabled (as in a pre-cancerous cell), the chromosome breakage syndrome may be tolerated and lead to accelerated neoplastic progression.

S-phase checkpoint function in itself (ref. 11; R.S. and D.M.L., unpublished work). This implies that BRCA1 plays a checkpoint-associated not a checkpoint-intrinsic function, possibly a direct role in SCR (Fig. 3). This is supported by the frequent "chromatid-type" errors that arise in *BRCA*-mutant MEFs^{11,16}.

Homozygous germline mutations in *ATM* produce genome instability and a cancer-prone syndrome in humans, and *Atm* participates in certain checkpoint responses to IR²¹. Indeed, *Atm* and a second checkpoint kinase, *Chk2*, signal to *BRCA1* after IR^{8,22}, but *Atm* does not appear to contribute to *BRCA1* phosphorylation after HU exposure^{5,8}. Moreover, unlike *ATR*^{-/-} embryos, *ATM*^{-/-} mice are viable. Taken together, the available data imply that the pathways that trigger *BRCA1* phosphorylation by these two proteins and the biological outcomes of these sets of events are, at least in part, distinct. Perhaps *Atr* and *Atm* activation are triggered by distinct categories of DNA structure.

Other genome integrity functions of the BRCA complex

Defective transcription-coupled repair of oxidative DNA damage was demonstrated in *BRCA1* and *BRCA2* mutated cells^{23,42}. Cells homozygous for a *BRCA1* allele lacking exon 11 revealed anomalous G2/M checkpoint responses to gamma irradiation, implying a role for *BRCA1* in this process⁹. Recently, a partial purification of *BRCA1* revealed its association with mismatch repair proteins, the Bloom's syndrome helicase, and certain replication factors²⁴. The functional significance of these particular associations is unclear. However, it is worth noting that the Bloom's Syndrome gene and certain mismatch repair genes have established tumour-suppressor function.

Unlike the homologues of the *RAD52* epistasis group of *S. cerevisiae*, which are represented in all eukaryotic genomes, there are no clear homologues of *BRCA* genes in yeast, flies or worms. Therefore, the biochemical contributions of *BRCA1* and *BRCA2* to DNA repair may be as regulators of more conserved repair functions.

Transcriptional functions of BRCA1 and BRCA2

Certain fragments of each *BRCA* protein, when fused to a GAL4 DNA-binding domain, can transactivate a GAL4 reporter, except when they bear clinically relevant mutations²⁵⁻²⁷. This suggested that *BRCA1* and/or *BRCA2* function as transcriptional regulators of specific target genes, the identification of which might, in turn, shed light on the mechanism of *BRCA* tumour suppression. Interactions have been found between *BRCA* proteins and sequence-specific transcription factors such as *c-myc* (*BRCA1*) and *p53* (*BRCA1* and *BRCA2*), as well as with certain transcription regulation proteins that do not recognize a specific, canonical DNA sequence (reviewed in refs 1 and 28). The search for *BRCA1* target genes has also pointed to candidates in the *p53* pathway: *p21* and *GADD45* (reviewed in refs 1 and 28, and references therein). This connection suggests a different kind of link between *BRCA1* function and genome integrity control in which its function is tied to the expression of genes that do the work of checkpoint control and/or DNA repair. *BRCA1*, when overproduced, also suppresses oestrogen receptor (ER) transactivation (cited in ref. 1). Notably, *p53* mutation and ER negativity are frequent findings in *BRCA1*-linked tumours, so *BRCA1* tumour-suppressor action appears not to be exclusively linked to *p53*- or *ER*-dependent functions.

Chromatin remodelling activity of BRCA multiprotein complexes

Chromatin remodelling functions have been attributed to both *BRCA1* and *BRCA2*^{1,28}. A recent biochemical purification of *BRCA1* suggests that it interacts with a SWI/SNF-containing complex²⁹. The impact of chromatin structure and of its remodelling upon transcriptional events is widely accepted. Equally important is the relationship between chromatin structure and the repair of recombinogenic DNA lesions. For example, in *S. cerevisiae*, the SIR proteins function in transcriptional silencing of subtelomeric

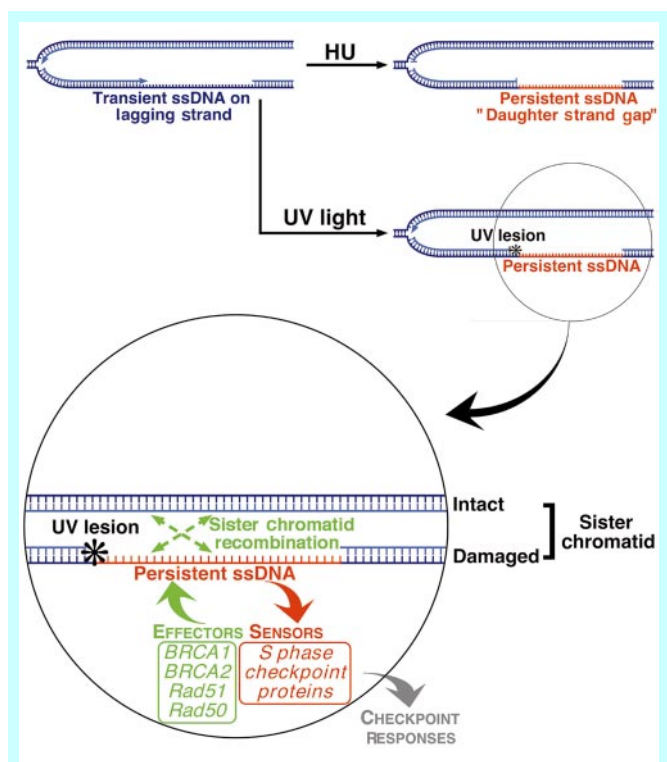


Figure 3 Proposed role for the *BRCA* proteins in sister chromatid recombination. DNA polymerase stalling activates S-phase checkpoint signalling and recruits *BRCA1*, *BRCA2*, *Rad51* and associated proteins to sites of arrested DNA synthesis. We propose that this recruitment arises from the existence of tracts of persistent ssDNA close to a replication fork. Such lesions, or their derivatives, may activate S-phase checkpoint signalling and act as a substrate for *BRCA* protein-mediated recombinational repair.

heterochromatin, but are rapidly released from such sites and recruited to the proximity of a DSB where their presence is required for efficient repair³⁰. This is an attractive paradigm against which to consider the *BRCA* proteins. Perhaps a DSB triggers the release of *BRCA* proteins from their 'civilian' functions or loci and promotes their recruitment to the 'locale' of the break.

If this concept is correct, what defines the 'locale' of a DSB? Recent studies have shown that chromatin is actively modified around a DSB in mammalian cells. Specifically, the mammalian histone H2A subspecies, H2AX, is phosphorylated on a residue in its C-terminal tail within minutes of gamma irradiation. Indeed, phosphorylated H2AX was shown to appear around an experimentally induced 'line' of DSBs in interphase nuclei (cited in ref. 31). Conceivably, phosphorylation of H2AX participates in defining the 'locale' of a DSB. Some phosphorylated H2AX also exists in undamaged cells, and it forms nuclear foci that co-localize, in part, with the *BRCA* dots³¹. This suggests that the *BRCA* dots may contain a specialized form of chromatin structure.

Although there are few clues to the innate biochemical activities of the *BRCA* gene products, it is notable that the *BRCA1* RING domain region, which interacts tightly, perhaps stoichiometrically, with the RING domain of *BARD1*, participates in an *in vitro* ubiquitin ligation reaction, like RING domains of some ubiquitin ligases³². Perhaps ubiquitination, sumoylation, or ned8 coupling of selected protein targets are mediated by *BRCA1* and/or *BARD1*.

BRCA genes and breast or ovarian cancer predisposition

There is good reason to accept that defective maintenance of genomic integrity, as described in *BRCA*-deficient cells, is an

accelerator of cancer progression. If so, how does such a generic defect translate into specifically increased breast or ovarian cancer risk?

The breast epithelium proliferates rapidly during puberty and under the influence of oestrogenic hormones. Unlike the cells of many rapidly proliferating epithelia, such as those of the intestine or of the uterine endometrium, progeny of this proliferative burst are retained within the breast epithelium. This is demonstrated by the finding that breast lobules are clonal³³. If, before lobular development, a lobular precursor cell had sustained a cancer-predisposing mutation at a relevant locus (such as *p53*), the entire lobule would then carry that mutation. This, in turn, could amplify the risk of subsequent neoplastic progression. Some germline *p53* mutations (such as in Li–Fraumeni syndrome) confer a particularly elevated risk of adult-onset breast cancer. Similarly, high-dose ionizing chest radiation during puberty, or even in early childhood (for example, atom-bomb exposure; mantle zone therapeutic radiation), specifically predisposes women to adult-onset breast cancer^{34,35}.

These observations could be relevant to *BRCA*-linked disease, especially if the *BRCA*^{+/-} genotype were haplo-insufficient in genome integrity maintenance function—an unanswered question at present. *BRCA* gene haplo-insufficiency could, in principle, increase the risk of additional cancer-promoting mutations occurring during breast development, including mutation of the *BRCA* gene, itself. Indeed, if puberty constitutes a limited ‘window’ during which a *BRCA* mutation carrier is at special risk of developing carcinogenic mutations, this might account for the relative absence of *BRCA* gene inactivation in sporadic breast or ovarian cancer. In sporadic disease, biallelic *BRCA* gene inactivation might occur too late to have an impact on disease risk³⁶.

That the breast and ovary are oestrogen-responsive tissues could also be relevant to the tissue specificity of *BRCA* disease risk. Some oestrogen metabolites can adduct DNA, and so could act as tissue-specific carcinogens (so-called “remote carcinogenesis”) ³⁷. Conceivably, this effect might be exacerbated by *BRCA* mutation, if the relevant DNA repair pathways were dysfunctional.

BRCA proteins associate with chromosomal pairing events on the synaptonemal complex. In some model organisms, homologous chromosomal pairing is also known to occur in certain somatic cells where it can influence development by affecting transcriptional regulation (“transvection”) as well as imprinting³⁸. Potentially analogous, homology-dependent transcriptional regulation may also operate in mammals³⁹ and could represent a link between homologous pairing and tissue-specific gene expression. If homologous chromosomal pairing occurs in the breast or ovarian epithelium, are *BRCA* proteins involved and does such an involvement contribute to their organ-specific tumour suppression function?

Human cells express alternatively spliced *BRCA1* transcripts, the biological significance of which is unclear⁴⁰. Whether products of such transcripts contribute to the tissue specificity of *BRCA* tumour suppression bears future investigation. □

- Welsh, P. L., Owens, K. N. & King, M. C. Insights into the functions of *BRCA1* and *BRCA2*. *Trends Genet.* **16**, 69–74 (2000).
- Scully, R. *et al.* Association of *BRCA1* with Rad51 in mitotic and meiotic cells. *Cell* **88**, 265–275 (1997).
- Sharan, S. K. *et al.* Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking *Brca2*. *Nature* **386**, 804–810 (1997).
- Chen, J. *et al.* Stable interaction between the products of the *BRCA1* and *BRCA2* tumour suppressor genes in mitotic and meiotic cells. *Mol. Cell* **2**, 317–328 (1998).
- Scully, R. *et al.* Dynamic changes of *BRCA1* subnuclear location and phosphorylation state are initiated by DNA damage. *Cell* **90**, 425–435 (1997).
- Wu, L. C. *et al.* Identification of a RING protein that can interact *in vivo* with the *BRCA1* gene product. *Nature Genet.* **14**, 430–440 (1996).
- Keegan, K. S. *et al.* The Atr and Atm protein kinases associate with different sites along meiotically pairing chromosomes. *Genes Dev.* **10**, 2423–2437 (1996).

- Cortez, D., Wang, Y., Qin, J. & Elledge, S. J. Requirement of ATM-dependent phosphorylation of *brca1* in the DNA damage response to double-strand breaks. *Science* **286**, 1162–1166 (1999).
- Deng, C. X. & Scott, F. Role of the tumor suppressor gene *Brca1* in genetic stability and mammary gland tumor formation. *Oncogene* **19**, 1059–1064 (2000).
- Hakem, R. *et al.* The tumor suppressor gene *Brca1* is required for embryonic cellular proliferation in the mouse. *Cell* **85**, 1009–1023 (1996).
- Patel, K. J. *et al.* Involvement of *Brca2* in DNA repair. *Mol. Cell* **1**, 347–357 (1998).
- Sonoda, E. *et al.* Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death. *EMBO J.* **17**, 598–608 (1998).
- Scully, R. *et al.* Genetic analysis of *BRCA1* function in a defined tumor cell line. *Mol. Cell* **4**, 1093–1099 (1999).
- Moynahan, M. E., Chiu, J. W., Koller, B. H. & Jasin, M. *Brca1* controls homology-directed DNA repair. *Mol. Cell* **4**, 511–518 (1999).
- Snowwaert, J. N. *et al.* *BRCA1* deficient embryonic stem cells display a decreased homologous recombination frequency and an increased frequency of non-homologous recombination that is corrected by expression of a *brca1* transgene. *Oncogene* **18**, 7900–7907 (1999).
- Scully, R., Puget, N. & Vlasakova, K. DNA polymerase stalling, sister chromatid recombination and the *BRCA* genes. *Oncogene* (in the press).
- Fornace, A. J. Jr Recombination of parent and daughter strand DNA after UV-irradiation in mammalian cells. *Nature* **304**, 552–554 (1983).
- Kowalczykowski, S. C. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* **25**, 156–165 (2000).
- Cordeiro-Stone, M., Makhov, A. M., Zaritskaya, L. S. & Griffith, J. D. Analysis of DNA replication forks encountering a pyrimidine dimer in the template to the leading strand. *J. Mol. Biol.* **289**, 1207–1218 (1999).
- Brown, E. J. & Baltimore, D. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Genes Dev.* **14**, 397–402 (2000).
- Rotman, G. & Shiloh, Y. ATM: a mediator of multiple responses to genotoxic stress. *Oncogene* **18**, 6135–6144 (1999).
- Lee, J. S., Collins, K. M., Brown, A. L., Lee, C. H. & Chung, J. H. hCds1-mediated phosphorylation of *BRCA1* regulates the DNA damage response. *Nature* **404**, 201–204 (2000).
- Gowen, L. C., Avrutskaya, A. V., Latour, A. M., Koller, B. H. & Leadon, S. A. *BRCA1* required for transcription-coupled repair of oxidative DNA damage. *Science* **281**, 1009–1012 (1998).
- Wang, Y. *et al.* BASC, a super complex of *BRCA1*-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev.* **14**, 927–939 (2000).
- Chapman, M. S. & Verma, I. M. Transcriptional activation by *BRCA1*. *Nature* **382**, 678–679 (1996).
- Monteiro, A. N., August, A. & Hanafusa, H. Evidence for a transcriptional activation function of *BRCA1* C-terminal region. *Proc. Natl Acad. Sci. USA* **93**, 13595–13599 (1996).
- Milner, J., Ponder, B., Hughes-Davies, L., Seltmann, M. & Kouzarides, T. Transcriptional activation functions in *BRCA2*. *Nature* **386**, 772–773 (1997).
- Irminger-Finger, I., Siegel, B. D. & Leung, W. C. The functions of breast cancer susceptibility gene 1 (*BRCA1*) product and its associated proteins. *Biol. Chem.* **380**, 117–128 (1999).
- Bochar, D. A. *et al.* *BRCA1* is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. *Cell* **102**, 257–265 (2000).
- Guarente, L. Diverse and dynamic functions of the Sir silencing complex. *Nature Genet.* **23**, 281–285 (1999).
- Paul, T. P. *et al.* A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr. Biol.* **10**, 886–895 (2000).
- Lorick, K. L. *et al.* RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc. Natl Acad. Sci. USA* **96**, 11364–11369 (1999).
- Kordon, E. C. & Smith, G. H. An entire functional mammary gland may comprise the progeny from a single cell. *Development* **125**, 1921–1930 (1998).
- Tokunaga, M. *et al.* Malignant breast tumors among atomic bomb survivors, Hiroshima and Nagasaki, 1950–74. *J. Natl Cancer Inst.* **62**, 1347–1359 (1979).
- Wolden, S. L., Lamborn, K. R., Cleary, S. F., Tate, D. J. & Donaldson, S. S. Second cancers following pediatric Hodgkin's disease. *J. Clin. Oncol.* **16**, 536–544 (1998).
- Kinzel, K. W. & Vogelstein, B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**, 761–763 (1997).
- Fishman, J., Osborne, M. P. & Telang, N. T. The role of estrogen in mammary carcinogenesis. *Ann. NY Acad. Sci.* **768**, 91–100 (1995).
- Henikoff, S. Nuclear organization and gene expression: homologous pairing and long-range interactions. *Curr. Opin. Cell Biol.* **9**, 388–395 (1997).
- Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. & Proudfoot, N. J. Intergenic transcription and transduction of the human beta-globin locus. *Genes Dev.* **11**, 2494–2509 (1997).
- Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266**, 66–71 (1994).
- Tibbetts, R. S. *et al.* Functional interactions between *BRCA1* and the checkpoint kinase ATR during genotoxic stress. *Genes Dev.* (in the press).
- Le Page, F. *et al.* *BRCA1* and *BRCA2* are necessary for the transcription-coupled repair of the oxidative 8-oxoguanine lesion in human cells. *Cancer Res.* **60**, 5548–5552 (2000).

Acknowledgements

We thank many colleagues for stimulating discussions and for sharing data before publication. In particular, J. Feunteun, R. Tibbetts, R. Abraham, W. Bonner, M. Gellert and P. Adams.

Correspondence and requests for materials should be addressed to D.M.L. (e-mail: david_livingston@dfci.harvard.edu).

The DNA damage response: putting checkpoints in perspective

Bin-Bing S. Zhou* & Stephen J. Elledge†

* Department of Oncology Research, SmithKline Beecham Pharmaceuticals, King of Prussia, Pennsylvania 19406, USA

† Department of Biochemistry and Molecular Biology, Department of Molecular and Human Genetics, Howard Hughes Medical Institute, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

The inability to repair DNA damage properly in mammals leads to various disorders and enhanced rates of tumour development. Organisms respond to chromosomal insults by activating a complex damage response pathway. This pathway regulates known responses such as cell-cycle arrest and apoptosis (programmed cell death), and has recently been shown to control additional processes including direct activation of DNA repair networks.

The survival of organisms depends on the accurate transmission of genetic information from one cell to its daughters. Such faithful transmission requires not only extreme accuracy in replication of DNA and precision in chromosome distribution, but also the ability to survive spontaneous and induced DNA damage while minimizing the number of heritable mutations. To achieve this fidelity, cells have evolved surveillance mechanisms that monitor the structure of chromosomes and coordinate repair and cell-cycle progression. Genetic control of cell-cycle transitions in response to DNA damage was first observed in the SOS DNA damage response pathway of *Escherichia coli*¹ and in mammals in ataxia telangiectasia (AT) cells, which are defective for the ataxia telangiectasia mutated (ATM) gene². This control was later observed in yeast and the term checkpoint was applied to the yeast pathway³.

DNA damage checkpoints were initially defined as non-essential regulatory pathways that control the ability of cells to arrest the cell cycle in response to DNA damage, allowing time for repair. However, recent evidence suggests that the historic definition of the checkpoint pathway is inadequate to explain the function of this pathway completely. In addition to controlling cell-cycle arrest, these pathways have been shown to control the activation of DNA repair pathways^{4–8}, the composition of telomeric chromatin and the movement of DNA repair proteins to sites of DNA damage^{9,10}, activation of transcriptional programmes¹¹, telomere length^{12,13} and, in some cell types for reasons not fully understood, induction of cell death by apoptosis^{14–17}. Thus, it is now clear that the checkpoint comprises a subroutine integrated into the larger DNA damage response pathway that regulates a multifaceted response (Fig. 1). In addition, several checkpoint genes are essential for cell and organism survival^{11,18–21}, implying that these pathways are not only surveyors of occasional damage, but are firmly integrated components of cellular physiology.

An example that illustrates the physiological importance of these pathways is the disorder ataxia telangiectasia. Individuals carrying two mutant ATM genes may suffer numerous problems that include loss of motor control owing to Purkinje cell loss, immune deficiencies and high frequencies of cancer. ATM is a central signalling protein in the DNA damage response, and so cells lacking ATM fail to execute many of the cellular responses to DNA damage. However, AT patients have many problems even in the absence of exposure to DNA-damaging agents because numerous chromosome structural defects occur during normal cell duplication. During the process of DNA replication, errors, such as double-strand breaks (DSBs) that arise from stalled replication forks, require attention by the DNA damage response pathway. In addition to these unavoidable errors, cells exist in a highly reactive chemical environment in which

reactive species, such as free radicals, are by-products of cellular metabolism or consequences of the cells' existence in an oxygen-rich environment. These reactive species react with DNA, creating more problems that must be addressed. In this regard, ATM has been proposed to be important in the cellular response to oxidative stress²², which may have particular relevance for long-lived, terminally differentiated cells such as Purkinje cells. Furthermore, DSBs are not solely accidents of nature, but are also regulated components of VDJ recombination and meiosis. ATM is needed to respond to these developmentally programmed DNA alterations as well, which emphasizes its importance to organism physiology.

The recently discovered connections between 'checkpoint' pathways and DNA repair and their physiological effects on the cell prompted us to re-evaluate the role of checkpoint proteins within the context of the overall response to DNA damage. Here we will concentrate primarily on data from mammalian cells, from which a model of the DNA damage response is beginning to take shape. There is much overlap between DNA damage response mechanisms

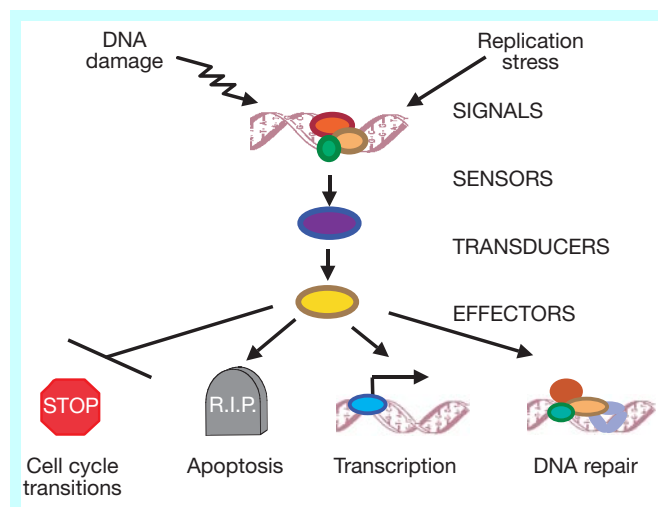


Figure 1 A contemporary view of the general outline of the DNA damage response signal-transduction pathway. Arrowheads represent activating events and perpendicular ends represent inhibitory events. Cell-cycle arrest is depicted with a stop sign, apoptosis with a tombstone. The DNA helix with an arrow represents damage-induced transcription, while the DNA helix with several oval-shaped subunits represents damage-induced repair. For the purpose of simplicity, the network of interacting pathways are depicted as a linear pathway consisting of signals, sensors, transducers and effectors.

in humans and yeast, and we will attempt to use these similarities to fill the gaps in our knowledge of the mammalian system.

General organization of the DNA damage response pathway

The DNA damage response pathway is a signal transduction pathway consisting of sensors, transducers and effectors (Fig. 1). Although we refer to this as a pathway, it is more accurately described as a network of interacting pathways that together execute the response. The identities of the sensors are not yet known. Much is known about the signal transducers, which are composed of four sets of conserved proteins with recognizable motifs (Table 1). One class is composed of phospho-inositide kinase (PIK)-related proteins which include ATM and ATM-Rad3-related (ATR) in mammals and their homologues in budding and fission yeast. These proteins are central to the entire DNA damage response. Downstream of these proteins are two families of checkpoint kinases (CHK), the Chk1 and Chk2 kinases, and their homologues. These kinases carry out subsets of the DNA damage response in mammals and are targets of regulation by ATM and ATR kinases. The fourth conserved family is the BRCT-repeat containing proteins, which include budding yeast Rad9 and fission yeast Crb2. No clear human orthologues of these proteins exist in mammals, but BRCA1 and 53BP1 are possible candidates. The precise roles of these proteins and where exactly they function with respect to the PIKs are not understood. However, *Saccharomyces cerevisiae* Rad9 is required for phosphorylation of budding yeast Chk1 and Chk2 homologues and interacts physically with both^{23–26}. For the Chk2 homologue, the interaction is dependent on the FHA domain, a phospho-amino-acid binding motif^{27,28}. Because PIKs are believed to phosphorylate Chk1 and Chk2 directly, it is possible that Rad9 and other BRCT proteins help in recruiting substrates like CHKs to PIKs in response to DNA damage²³.

Below this level of signal transduction are the effectors that execute the functions of the DNA damage response. These include substrates of both PIK and CHK kinases and proteins involved in DNA repair, transcription regulation and cell-cycle control, such as BRCA1, Nbs1, p53 and Cdc25C (Table 2). Depending on the context, certain molecules such as BRCA1 may have multiple functions in this signal transduction pathway.

DNA damage sensors

The proteins that initially sense the aberrant DNA structures, and initiate the signalling response, are currently unknown. Owing to their ability to bind and be activated by DNA strand breaks, poly (ADP-ribose) polymerase (PARP) and DNA-dependent protein kinase (DNA-PK) have long been proposed as DNA damage sensors. However, genetic evidence indicates that these proteins are not activators of the global DNA damage response^{29,30}. In place of these molecules, a group of four conserved proteins in yeast have emerged as candidates because they share some of the properties

expected for sensors, including an essential genetic role in the activation of the DNA damage response pathway and the potential to interact with nucleic acids. In *Schizosaccharomyces pombe*, three of these proteins, Rad1, Rad9 and Hus1, are related in structure to PCNA (Table 1). Modelling suggests that they may form a doughnut-like heteromer like PCNA, and in principle could be loaded onto damaged DNA just as PCNA is loaded onto primed DNA (reviewed in ref. 31). The human homologues of these proteins form a DNA damage responsive complex³². Consistent with yeast studies, inactivation of mouse Hus1 results in impaired responses to genotoxic stress³³.

The fourth conserved checkpoint protein from *S. pombe*, Rad17, shares homology with all five subunits of replication factor C (RFC)³¹. During replication, the RFC1–5 complex binds to DNA primed by primase and DNA pol- α , and recruits PCNA to form a sliding clamp that tethers DNA polymerases to the DNA template. This complex could signal to maintain the replication checkpoint during S phase and, in fact, mutations in RFCs have been found to be defective for the DNA replication checkpoint^{34–37}. Analogously, in the DNA damage response, Rad17 homologues, together with the small RFC subunits, can mediate interactions with the Rad1–Rad9–Hus1 complex³⁸, and are possibly involved in maintaining the damage signal until repair is completed.

The simplest model concerning these putative sensors is that Rad17 complexes bind to damaged DNA directly, and then load the Rad1 complexes onto DNA, which leads to activation of the main damage response kinase Rad3 in *S. pombe*. But in *S. pombe*, although the known biochemical outputs of this damage response, including Rad3-dependent Cds1 and Chk1 phosphorylation, are dependent on Rad1 and Rad17 complexes, Rad3-dependent phosphorylation of Rad26 in response to ionizing radiation (IR) does not require these proteins³⁹. Thus, these putative sensors are not always required for Rad3-dependent phosphorylation events. Similar phosphorylation dependency after DNA damage was also observed for a *S. cerevisiae* Rad26 homologue, Ddc2/Lcd1 (refs 40, 41), suggesting that a human homologue could exist and behave in a similar manner (Table 1). But the phosphorylation of Rad26 does not necessarily exclude a sensory role for the Rad1 and Rad17 complexes. It is possible that Rad26 or Rad3 is able to recognize a DSB leading to phosphorylation of the tightly bound Rad26 protein, but full activation of the Rad3 pathway requires the function of the Rad1 and Rad17 complexes to amplify the damage signal. In support of this, Rad26 phosphorylation in response to replication blocks is dependent on the Rad1 complex of proteins³⁹. Alternatively, it is possible that there are two separate sensory pathways, each of which must function to fully activate the damage response. This may provide a failsafe mechanism to ensure the pathway is not activated spuriously to arrest the cell cycle inappropriately or activate cell death.

If the above proteins are not the actual sensors, why have the identities of the true sensors not yet emerged from the powerful genetic studies in yeast? One possibility is that we have insufficient biochemical information to recognize them as such. Or there may be many different sensors that have overlapping abilities to sense DNA damage and to signal. This situation is further complicated by the fact that a given DNA damaging agent generates several types of DNA damage, possibly activating several sensors.

Alternative proteins have been suggested as candidate sensors, one of which is the breast cancer protein BRCA1. Mouse cells lacking BRCA1 exon 11 are unable to arrest the cell cycle in G2, suggesting a possible defect in sensing or signalling⁴². BRCA1 is part of a large complex named BASC (BRCA1-associated genome surveillance complex) that contains ATM, the Nbs1–MRE11–RAD50 complex, mismatch proteins (MSH2/6 and MLH2), and the Bloom's helicase (BLM)⁴³. This model is speculative, but it is interesting that each of these proteins has the ability to recognize aberrant DNA structures and could thus be involved in transmitting

Table 1 Conserved DNA damage response genes

Functional class	Mammals	<i>S. cerevisiae</i>	<i>S. pombe</i>
PCNA-like proteins	Rad1 Rad9 Hus1	Rad17 Ddc1 Mec3	Rad1 Rad9 Hus1
RFC-like proteins	Rad17 RFC2–5	Rad24 RFC2–5	Rad17 RFC3
BRCT proteins	BRCA1?, 53BP1?	Rad9 DPB11	Crb2/Rhp9 Cut5
P13K-like proteins	ATR ATM	Mec1 Tel1	Rad3 Tel1
Effector kinases	Chk1 Chk2	Chk1 Rad53	Chk1 Cds1
Coiled-coil proteins	?	Ddc2/Lcd1	Rad26

the presence of these structures to ATM and BRCA1. Initial support for this idea is provided by the observation that cells defective for the mismatch repair gene *MLH1* fail to display the ATM-dependent phosphorylation of c-Abl in response to *cis*-platinum treatment⁴⁴. It remains to be seen whether *MLH1* mutants are defective in all responses to *cis*-platinum, for example, in cell-cycle arrest. Interestingly, several of these proteins are also substrates of ATM, suggesting that they may be targets of regulation as opposed to directing signal transduction. At present, it is impossible to distinguish between a sensory role and an effector role for any of these proteins, and it is plausible that the activity of an entire complex must be intact to properly sense and respond to damage.

Proximal kinases ATM and ATR

In contrast to our knowledge of damage sensors, our understanding of signal transducers is more advanced. Two related and conserved proteins, ATM and ATR, are central components of the DNA damage response¹¹. Although structurally related to the PI(3)K family members, ATM and ATR are protein kinases. The function of ATM is better understood. Cells from AT patients have mutations in ATM and are defective in several responses to IR including G1 arrest⁴⁵, reduction in DNA synthesis² and G2 arrest⁴⁶. ATM plays an important part in the response to IR, controlling the initial phosphorylation of several key proteins such as p53, Mdm2, BRCA1, Chk2 and Nbs1 in response to DNA damage (Fig. 2). These proteins are still phosphorylated in γ -irradiated AT cells, but with delayed kinetics, indicating that additional pathways respond to IR. Whereas AT fibroblasts are very sensitive to IR, they show little sensitivity to ultraviolet radiation, alkylating agents or inhibitors of DNA replication.

Less is known about ATR owing to the absence of ATR mutations in human diseases. However, expression of a dominant-negative ATR sensitizes mammalian cells to all forms of DNA damage and diminishes the G2/M checkpoint response induced by γ -radiation^{47,48}. Thus, ATR is a good candidate for a pathway that operates in parallel to ATM and responds to different types of damage. Additional support for this model comes from the observation that *S. cerevisiae* ATM and ATR homologues have overlapping but non-redundant roles¹¹. Furthermore, ATM and ATR share a number of phosphorylation substrates and often phos-

phorylate the same residues, although substrate differences have been reported⁴⁹. One substrate that has been examined, p53, is phosphorylated at Ser 15 by both kinases^{50–52} *in vitro* and *in vivo*. In addition, ATM and ATR can phosphorylate BRCA1 at Ser 1423, 1524 and other sites^{4,53}. BRCA1 phosphorylation in response to IR is largely ATM-dependent but phosphorylation in response to ultraviolet light and hydroxyurea requires ATR⁵³. These data fit the emerging model that ATM primarily controls the response to IR, whereas ATR responds to other types of damage. One caveat in the interpretation of the *in vivo* roles of ATR is that most experiments have been performed with a dominant-negative ATR construct that might also interfere with ATM function. Resolution of this will require viable loss-of-function ATR mutants.

The phenotypes of knockout mice have been particularly revealing concerning the function of PIKs. While ATM null mice are viable and display growth retardation and infertility^{16,54,55}, *ATR*^{−/−} mice die early in embryogenesis^{18,19}. *ATR*^{−/−} blastocyst cells die in culture with a phenotype resembling mitotic catastrophe¹⁸. The latter result suggests an essential role for ATR, possibly monitoring DNA replication. The *S. cerevisiae* ATR homologue, Mec1, is also essential and has been shown to be required for recovery from DNA replicational stress⁵⁶. Thus, ATR is likely to have a role in each cell cycle, indicating either a constitutive function for ATR or the frequent presence of spontaneous DNA damage or replicational stress during a normal cell cycle.

Precisely how these kinases are controlled in response to various stimuli is unknown. ATM kinase activity can be activated by DNA damage *in vivo*^{50,51,57}. Direct activation by DNA is less well established. Some studies report that there is no stimulation by DNA *in vitro*^{51,58}; others have found that small amounts of purified ATM bind to and are activated by DNA with DSBs *in vitro*⁵⁹. It is not clear, however, whether ATM alone retains the ability to bind DNA or whether other proteins complexed with it are required. ATM may ultimately resemble DNA-PK, which by itself has a low affinity for DSB-containing DNA, but which acquires a significantly enhanced affinity in the presence of a DNA end-binding factor, the Ku 70/80 complex, which loads DNA-PK onto DNA. Proteins to which ATM is bound are candidate sensors whose identification will be essential for the molecular explanation of ATM activation and DNA damage recognition. Similar activation by damage has not been observed for ATR. However, ATR is likely to be regulated in some fashion because it controls the late phosphorylation of p53 in response to IR^{52,60,61} and the ultraviolet-induced phosphorylation of Chk1 (ref. 20). Thus, it is likely that ATR activation is different from that of ATM and may involve a type of substrate accessibility activation that is not readily recapitulated *in vitro*. For example, sensor proteins may become activated and recruit substrates to ATR. This type of mechanism has been suggested to explain the observation that *S. cerevisiae* Rad9, a BRCT-repeat-containing protein, is required for Rad53 phosphorylation in response to DNA damage²³. Rad9 is phosphorylated in response to DNA damage and phosphorylated Rad9 binds Rad53. A substrate recruitment mechanism has the additional advantage of allowing particular damage sensors to recruit distinct subsets of substrates to tailor the response to each type of damage. Recently, ATR was found to respond to DNA damage and replication blocks by forming foci that overlap with BRCA1 foci⁵³. The foci formed in response to replication blocks are at sites of ongoing DNA replication. Thus, it is possible that ATR is regulated and localized to its substrates in response to genotoxic stress. Unravelling ATR regulation promises to be of primary importance in understanding the DNA damage response pathway.

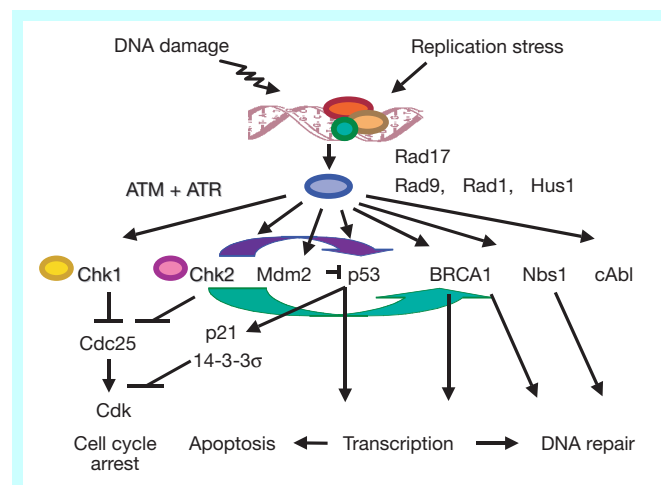


Figure 2 Organization of the mammalian DNA damage response pathway. Arrowheads represent positively acting steps while perpendicular ends represents inhibitory steps. Gene names are shown at the approximate positions where their encoded proteins function in the pathway. Although the general organization of the pathway is correct, some details are omitted, especially concerning the relationship between the ATR/ATM and Hus1/Rad17/Rad9/Rad1 proteins, which may participate in mutual regulation.

Downstream kinases Chk1 and Chk2

How do ATM and ATR control downstream DNA damage responses? Current evidence suggests it is partially by controlling Chk1 and Chk2 kinases: two serine/threonine kinases that are structurally unrelated but that share some overlapping substrate

specificity. Chk2 is the homologue of yeast Rad53 and Cds1 (Table 1) that are required for responses to DNA damage and replication blocks, whereas Chk1 kinases in yeast are primarily responsible for cell-cycle arrest in response to DNA damage. Chk2/hCds1 is phosphorylated and activated in response to IR, methyl methane sulphonate and hydroxyurea^{62–65}. Its phosphorylation in response to IR is ATM-dependent. It contains a phospho-amino acid binding motif called the FHA domain and a SQ cluster domain (SCD) in its amino terminus. Thr 68 in the SCD is phosphorylated by ATM *in vivo* and *in vitro* in response to damage^{66–68}. Chk2 phosphorylation is required for its activation^{62,67,68}. ATR also phosphorylates these sites and may regulate the responses to ultraviolet and other agents. *CHK2* null mutant ES cells have been shown to be defective in maintaining but not initiating G2 arrest in response to IR. More importantly, IR-treated *CHK2*^{−/−} thymocytes fail to stabilize and activate p53, to induce p53-dependent inducible genes and to trigger apoptosis¹⁷. Their inability to activate p53 and p21 indicates that *CHK2* mutants are not able to arrest in G1 in response to IR. Dominant-negative *CHK2* mutants also prevent p53 activation⁶⁹. Chk2 can phosphorylate p53 on Ser 20 *in vitro*^{17,69,70}, which has been shown to interfere with binding to the ubiquitin ligase Mdm2 (ref. 71). Other studies, however, suggest that Ser 20 phosphorylation is not important for p53 stabilization in certain contexts such as ultraviolet damage⁷². Furthermore, ATM phosphorylates Mdm2 (ref. 73), the significance of which is unknown. Thus, the mechanism of ATM and Chk2 regulation of p53 is still controversial. As ATM and Chk2 both phosphorylate p53, they may synergize to ensure p53 activation only when the pathway is fully active. Genetic evidence for a common pathway for Chk2 and p53 is also provided by the identification of mutations in the *CHK2* gene in a subset of Li–Fraumeni syndrome cancer families that lack p53 mutations⁷⁴. Chk2 can phosphorylate BRCA1 (ref. 75), indicating yet more roles for Chk2.

Unlike Chk2, Chk1 kinase activity does not appear to be increased in response to DNA damage or replication blocks. Nevertheless, Chk1 is phosphorylated in response to IR in mammals^{20,76} and in yeast^{26,77}. Phospho-specific antibodies recognizing phosphoserine 345 reveal that Chk1 is dramatically phosphorylated on Ser 345 in response to hydroxyurea and ultraviolet light, but only moderately phosphorylated in response to IR. Ser 345 phosphorylation in response to ultraviolet light is ATR-dependent *in vivo*²⁰ and ATR can phosphorylate this site *in vitro*⁴⁹. The role of Chk1 phosphorylation is currently unknown, but it is likely to facilitate Chk1 function. As the kinase activity of phosphorylated Chk1 is not highly activated toward general substrates *in vitro*, phosphorylation may permit it to assemble with its relevant substrates *in vivo* to execute its function. Mice lacking *CHK1* die in early embryogenesis similarly to *ATR*^{−/−} mice^{20,21}. The fact that *ATR* and *CHK1* mutant mice have similar phenotypes and that ATR controls Chk1 phosphorylation suggests that Chk1 is a key effector of the ATR pathway and may be responsible for its essential function.

Studies on Chk1 function have progressed slowly because *CHK1* is an essential gene. Embryonic stem cells expressing a conditional *CHK1* gene die of p53-independent apoptosis after loss of *CHK1*. Before their death, these cells become incapable of preventing mitotic entry in response to IR²⁰, demonstrating that Chk1 is required for the G2 DNA damage checkpoint in mammals, as previously observed in other organisms¹¹. The fact that Chk1 is essential for cell viability suggests that it also has important functions during S phase, possibly facilitating DNA replication and preventing premature mitotic entry.

How do Chk1 and Chk2 prevent mitotic entry? Arrest in G2 is mainly regulated by the maintenance of inhibitory phosphorylation of Cdc2 (ref. 78). Cdc2 dephosphorylation and activation is catalysed by the dual specificity phosphatase Cdc25 (ref. 79). Recent evidence indicates that part of the G2/M DNA checkpoint mechanism involves inactivation and translocation of Cdc25C into

the cytoplasm. This is at least partially mediated by phosphorylation on Ser 216 in Cdc25C and its consequent binding with 14-3-3 proteins^{80–82}. Chk1 (ref. 76) and Chk2 (refs 62, 64, 65) have been shown to phosphorylate Cdc25C at Ser 216 *in vitro*. This modification is thought to maintain Cdc25C phosphorylation in cells arrested at G2/M in response to DNA damage. Genetic evidence supports this model, but biochemical confirmation using cultured mammalian cells has not yet been achieved, possibly because Cdc25C is normally phosphorylated on Ser 216 during S phase and DNA damage merely prolongs its phosphorylation. Because p53 targets such as p21 and 14-3-3σ have roles in maintaining G2 arrest^{83,84}, Chk2 regulation of p53 is also expected to be important for the G2 DNA damage checkpoint. It is likely that other components of Cdc2 regulation are controlled in response to DNA damage, such as the inhibitory Wee1 and Myt1 kinases and cyclin B localization, all of which are known to regulate mitotic entry.

DNA damage response and repair are interacting networks

Recent observations have made it clear that the response to DNA damage in mammals is not limited to decisions on cell-cycle arrest and apoptosis, but is intimately involved in controlling repair itself. DNA repair pathways consist of an intricate network of repair systems that each target a specific subset of lesions. Much of DNA repair is constitutive, but a number of regulatory connections between the DNA damage response pathway and DNA repair have emerged (Fig. 3). First, in yeast and mammals, a large number of genes involved in DNA repair are transcriptionally induced in response to DNA damage in a DNA damage response pathway dependent manner, suggesting that many facets of repair are enhanced^{11,85}. Second, fibroblasts lacking p53 have been shown to be defective in global excision repair of cyclobutane dimers. The p48 gene, which is mutated in *Xeroderma pigmentosum* group E cells, is induced by DNA damage in a p53-dependent fashion⁸⁶, possibly explaining p53's role in excision repair. Recently, a new nuclear localized subunit of ribonucleotide reductase, p53R2, was found to be induced by p53 in response to DNA damage. Blocking p53R2 expression increases cell killing by a variety of DNA-damaging agents⁸⁷, supporting a functional role for p53R2 in DNA repair. Regulation of ribonucleotide reductase through the DNA damage response pathway represents a conserved strategy employed by the DNA damage response kinases to facilitate repair¹¹. Together these results challenge the long-held notion that p53 functions mainly to induce apoptosis and suggest that p53 also

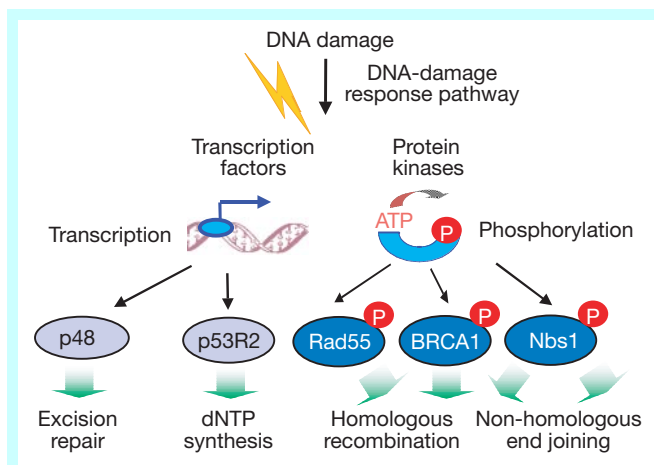


Figure 3 Interactions between the DNA damage response pathway and DNA repair networks. The regulatory connections include transcriptional upregulation of repair proteins such as p48 and p53R2 after DNA damage, and phosphorylation of repair proteins such as BRCA1, Nbs1 and Rad55 after DNA damage. See text for details.

promotes cell survival in response to DNA damage under certain circumstances.

Another example of regulation of DNA repair comes through studies of Nijmegen breakage syndrome (NBS), which has been classified as a variant of AT on the basis of similarities between the two diseases. The Nbs1 protein resides, together with hMre11 and hRad50 (ref. 88), in a complex that is required for both non-homologous end joining (NHEJ) and homologous recombination repair of DSBs in DNA. Recently, hMre11 was also found to be mutated in individuals with another AT-like disorder⁸⁹. AT cells have a reduced capacity to rejoin breaks⁹⁰. Moreover, both ATM and Nbs1 are required for protection against radio-resistant DNA synthesis. These similarities suggest that ATM and the Nbs1–hMre11–Rad50 complex function in the same pathway. Further evidence has come from recent observations that ATM directly phosphorylates Nbs1 on several sites that are needed for its function *in vivo*^{5–8}. This connection helps to explain how ATM regulates recombination and repair in response to DNA damage. In addition, both ATM and the Nbs1–Mre11–Rad50 complex are components of a large complex of proteins, BASC, which contains BRCA1. BRCA1 is also an ATM substrate and is required for DNA damage-induced homologous recombination⁹¹, which suggests that these proteins are all regulated by ATM to execute repair processes (Fig. 2). Finally, the yeast Rad55 recombination protein⁹² and the single-strand DNA binding protein RPA⁹³ are phosphorylated in response to DNA damage, further demonstrating that DNA repair proteins are regulated directly by the DNA damage response pathway. It is likely that identification of other repair proteins as substrates of damage response kinases will soon follow.

Another apparent connection to DNA repair comes from experiments in *S. cerevisiae* that show that *rad9* and *rad17* mutants have opposite effects on the rate of excision of telomeric DNA when the telomeres are no longer protected by telomeric binding proteins⁹⁴. How they regulate these repair activities and why they behave in opposing fashions are not understood. Nevertheless, these proteins clearly play a role in controlling DNA repair functions, giving rise to the possibility that these proteins initially functioned in DNA repair and later evolved into signalling molecules.

Summary and future directions

During the past several years, considerable progress has been made in elucidating the components of the eukaryotic DNA damage response. It has become clear that the checkpoint pathways, originally thought to control primarily cell-cycle decisions, are actually part of a vast DNA damage response pathway, only a branch of which is involved in cell-cycle control (Fig. 1). Thus, for purposes of clarity it is best to refer to the entire pathway as the DNA damage response and to those components specifically involved in controlling cell-cycle progression as the 'checkpoint' branch.

How damaged DNA activates DNA damage response proteins is a central question to be addressed in the next few years. In principle, a sensor protein should have the ability to interact with damaged DNA. There are many ways to damage DNA. Thus, it will be important to determine whether there must be a sensor for each type of damage, or, alternatively, whether each type of damage is converted into one of a few common intermediates (for example, single-strand DNA), that can be detected by a limited number of sensor proteins. In addition, the process of sensing sets up a potential conflict for the cell. If damage itself is sensed through interaction with a sensor protein, this sensor protein may potentially interfere with repair proteins that also need access to damaged DNA. How does the cell resolve this potential conflict? A corollary issue is whether it is the damaged DNA or the process of repair that is sensed. The cell must be aware not only of damage, but also of when the damage is repaired, because completion of repair should be the signal for termination of the response and resumption of the cell cycle. *E. coli* has solved this problem in a parsimonious fashion

because its 'sensor' is the single-strand DNA binding protein RecA, the recombinase involved in repairing the damage. Thus, in *E. coli*, it is repair that is monitored. Whether eukaryotes will once again turn out to mimic *E. coli* remains to be seen.

How the damage-signalling kinases are activated and select substrates are critical unresolved issues. Both ATM and Chk2 are activated in response to damage, but ATR and Chk1 regulation remains enigmatic. Furthermore, although many kinase substrates have been identified (Table 2), detailed mechanistic characterization of these substrates and identification of new substrates will be needed to unravel the vast complexities of the DNA damage response.

Why must the cell regulate repair? A plausible explanation is that certain constitutive repair capacities can be deleterious in certain situations. Optimal repair of various types of damage might require the regulated presence of certain repair mechanisms and the absence of others. For example, cells might benefit from a mechanism that favours homologous recombination for DSB repair in sister chromatids as opposed to non-homologous end joining. Alternatively, during DNA replication some structures produced transiently at replication forks may resemble damage, and inappropriate attempts to repair them by constitutive repair functions could disrupt DNA synthesis. This may be in part why cells stop initiation of DNA synthesis globally in the presence of DNA damage while repair is initiated. In addition, repair is unlikely to consist solely of manipulating DNA. If DNA polymerase encounters damage, replication forks may need to be reorganized to facilitate repair, possibly by restructuring chromatin to allow accessibility to repair proteins. Restructuring of chromatin may similarly be required at sites of breaks and would be consistent with the recruitment of Ku and Sirs to sites of DSBs. BRCA1 was also recently shown to be part of a chromatin remodelling complex⁹⁵. A chromatin remodelling capacity of this sort would require tight regulation to avoid interference with ongoing cellular processes.

Another area ripe for future exploration is the relationship between the DNA damage signalling pathway and telomeric DNA structures⁹⁶. It has been proposed that critically shortened telomeres activate the DNA damage response pathway, thereby preventing genomic instability and cancer in ageing cells. Conversely, mutants in ATM and TEL1 result in shortened telomeres and display some aspects of premature ageing. The relationships between ageing, damage response pathways and telomeres need to be better understood and integrated into models for tumorigenesis.

Thus, the pathways through which cells sense and transduce DNA damage signals have begun to come into focus. It is now clear that, although these pathways control cell-cycle transitions and apoptotic decisions, its overall goal is the optimal repair of DNA damage such

Table 2 Targets of damage response kinase-dependent phosphorylation

Kinase class	Kinase substrate		
	Mammals	<i>S. cerevisiae</i>	<i>S. pombe</i>
PIK homologues	p53*, Mdm2*		
ATM, ATR (Mammals)	Chk1*, Chk2*	Rad53*, Rad9‡	Chk1‡, Cds1‡
Mec1, Tet1 (<i>S. cerevisiae</i>)	BRCA1*, CtIP*	Ddc1‡, Ddc2‡	Rad26‡, Hus1‡
Rad3, Tel1 (<i>S. pombe</i>)	Nbs1*, c-Abl*	Rpa2‡	Crb2‡
	RPA-p34*		
Chk1 homologues			
Chk1 (Mammals)	Cdc25C†		
Chk1 (<i>S. cerevisiae</i>)	Cdc25A†	Pds1‡	Cdc25†
Chk1 (<i>S. pombe</i>)	p53*		Wee1†
Chk2 homologues			
Chk2 (Mammals)	Cdc25C†	Ort1‡, Dun1‡	
Rad53 (<i>S. cerevisiae</i>)	p53*	Rad55‡, Swi6‡	Cdc25†
Cds1 (<i>S. pombe</i>)	BRCA1†	Dbf4‡, Cdc5‡	Wee1†

* *In vitro* phosphorylation and *in vivo* dependency.

† *In vitro* phosphorylation.

‡ *In vivo* dependency.

that cells can go on to function and contribute to the long-term survival of the organism and species with minimal levels of genetic variation. The next decade promises to be rich with new insights into this system. □

1. George, J., Castellazzi, M. & Buttin, G. Prophage induction and cell division in *E. coli*. III. Mutations *sfiA* and *sfiB* restore division in *tif* and *lon* strains and permit the expression of mutator properties of *tif*. *Mol. Gen. Genet.* **140**, 309–332 (1975).
2. Painter, R. B. & Young, B. R. Radiosensitivity in ataxia-telangiectasia: A new explanation. *Proc. Natl. Acad. Sci. USA* **77**, 7315–7317 (1980).
3. Weinert, T. A. & Hartwell, L. H. The Rad9 gene controls the cell cycle response to DNA damage in *Saccharomyces cerevisiae*. *Science* **241**, 317–322 (1988).
4. Cortez, D., Wang, Y., Qin, J. & Elledge, S. J. Requirement of ATM-dependent phosphorylation of *brca1* in the DNA damage response to double-strand breaks. *Science* **286**, 1162–1166 (1999).
5. Lim, D. S. *et al.* ATM phosphorylates p95/nbs1 in an S-phase checkpoint pathway. *Nature* **404**, 613–617 (2000).
6. Gatei, M. *et al.* ATM-dependent phosphorylation of nibrin in response to radiation exposure. *Nature Genet.* **25**, 115–119 (2000).
7. Zhao, S. *et al.* Functional link between ataxia-telangiectasia and Nijmegen breakage syndrome gene products. *Nature* **405**, 473–477 (2000).
8. Wu, X. *et al.* ATM phosphorylation of Nijmegen breakage syndrome protein is required in a DNA damage response. *Nature* **405**, 477–482 (2000).
9. Martin, S. G., Laroche, T., Suka, N., Grunstein, M. & Gasser, S. M. Relocalization of telomeric Ku and SIR proteins in response to DNA strand breaks in yeast. *Cell* **97**, 621–633 (1999).
10. Mills, K. D., Sinclair, D. A. & Guarente, L. MEC1-dependent redistribution of the Sir3 silencing protein from telomeres to DNA double-strand breaks. *Cell* **97**, 609–620 (1999).
11. Elledge, S. J. Cell cycle checkpoints: preventing an identity crisis. *Science* **274**, 1664–1672 (1996).
12. Naito, T., Matsuura, A. & Ishikawa, F. Circular chromosome formation in a fission yeast mutant defective in two ATM homologues. *Nature Genet.* **20**, 203–206 (1998).
13. Ritchie, K. B., Mallory, J. C. & Petes, T. D. Interactions of TLC1 (which encodes the RNA subunit of telomerase), TEL1, and MEC1 in regulating telomere length in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**, 6065–6075 (1999).
14. Lowe, S. W., Schmitt, E. M., Smith, S. W., Osborne, B. A. & Jacks, T. p53 is required for radiation-induced apoptosis in mouse thymocytes. *Nature* **362**, 847–849 (1993).
15. Clarke, A. R. *et al.* Thymocyte apoptosis induced by p53-dependent and independent pathways. *Nature* **362**, 849–852 (1993).
16. Xu, Y. & Baltimore, D. Dual roles of ATM in the cellular response to radiation and in cell growth control. *Genes Dev.* **10**, 2401–2410 (1996).
17. Hirao, A. *et al.* DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science* **287**, 1824–1827 (2000).
18. Brown, E. J. & Baltimore, D. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Genes Dev.* **14**, 397–402 (2000).
19. de Klein, A. *et al.* Targeted disruption of the cell-cycle checkpoint gene ATR leads to early embryonic lethality in mice. *Curr. Biol.* **10**, 479–482 (2000).
20. Liu, Q. *et al.* Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. *Genes Dev.* **14**, 1448–1459 (2000).
21. Takai, H. *et al.* Aberrant cell cycle checkpoint function and early embryonic death in Chk1 (-/-) mice. *Genes Dev.* **14**, 1439–1447 (2000).
22. Rotman, G. & Shiloh, Y. Ataxia-telangiectasia: is ATM a sensor of oxidative damage and stress? *BioEssays* **19**, 911–917 (1997).
23. Sun, Z., Hsiao, J., Fay, D. S. & Stern, D. F. Rad53 FHA domain associated with phosphorylated Rad9 in the DNA damage checkpoint. *Science* **281**, 272–274 (1998).
24. Emili, A. MEC1-dependent phosphorylation of Rad9p in response to DNA damage. *Mol. Cell* **2**, 183–189 (1998).
25. Vialard, J. E., Gilbert, C. S., Green, C. M. & Lowndes, N. F. The budding yeast Rad9 checkpoint protein is subjected to Mec1/Tel1-dependent hyperphosphorylation and interacts with Rad53 after DNA damage. *EMBO J.* **17**, 5678–5688 (1998).
26. Sanchez, Y. *et al.* Control of the DNA damage checkpoint by chk1 and rad53 protein kinases through distinct mechanisms. *Science* **286**, 1166–1171 (1999).
27. Durocher, D., Henkel, J., Fersht, A. R. & Jackson, S. P. The FHA domain is a modular phosphopeptide recognition motif. *Mol. Cell* **4**, 387–394 (1999).
28. Liao, H., Byeon, I. J. & Tsai, M. D. Structure and function of a new phosphopeptide-binding domain containing the FHA2 of Rad53. *J. Mol. Biol.* **294**, 1041–1049 (1999).
29. Wang, Z. Q. *et al.* Mice lacking ADPRT and poly(ADP-ribosyl)ation develop normally but are susceptible to skin disease. *Genes Dev.* **9**, 509–520 (1995).
30. Jimenez, G. S. *et al.* DNA-dependent protein kinase is not required for the p53-dependent response to DNA damage. *Nature* **400**, 81–83 (1999).
31. O'Connell, M. J., Walworth, N. C. & Carr, A. M. The G2-phase DNA damage checkpoint. *Trends Cell Biol.* **10**, 296–303 (2000).
32. Volkmer, E. & Karnitz, L. M. Human homologs of *Schizosaccharomyces pombe* rad1, hus1, and rad9 form a DNA damage-response protein complex. *J. Biol. Chem.* **274**, 567–570 (1999).
33. Weiss, R. S., Enoch, T. & Leder, P. Inactivation of mouse hus1 results in genomic instability and impaired responses to genotoxic stress. *Genes Dev.* **14**, 1886–1898 (2000).
34. Sugimoto, K., Ando, S., Shimomura, T. & Matsumoto, K. Rfc5, a replication factor C component, is required for regulation of Rad53 protein kinase in the yeast checkpoint pathway. *Mol. Cell. Biol.* **17**, 5905–5914 (1996).
35. Noskov, V. N., Araki, H. & Sugino, A. The RFC2 gene, encoding the third-largest subunit of the replication factor C complex, is required for an S-phase checkpoint in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **18**, 4914–4923 (1998).
36. Shimomura, T., Ando, S., Matsumoto, K. & Sugimoto, K. Functional and physical interaction between Rad24 and Rfc5 in the yeast checkpoint pathways. *Mol. Cell. Biol.* **18**, 5485–5491 (1998).
37. Shimada, M. *et al.* Replication factor C3 of *Schizosaccharomyces pombe*, a small subunit of replication factor C complex, plays a role in both replication and damage checkpoints. *Mol. Biol. Cell* **10**, 3991–4003 (1999).
38. Green, C. M., Erdjument-Bromage, H., Tempst, P. & Lowndes, N. F. A novel Rad24 checkpoint protein complex closely related to replication factor C. *Curr. Biol.* **10**, 39–42 (2000).
39. Edwards, R. J., Bentley, N. J. & Carr, A. M. A Rad3-Rad26 complex responds to DNA damage independently of other checkpoint proteins. *Nature Cell Biol.* **1**, 393–398 (1999).
40. Paciotti, V., Clerici, M., Lucchini, G. & Longhese, M. P. The checkpoint protein Ddc2, functionally related to *S. pombe* Rad26, interacts with Mec1 and is regulated by Mec1-dependent phosphorylation in budding yeast. *Genes Dev.* **14**, 2046–2059 (2000).
41. Rouse, J. & Jackson, S. P. LCD1: An essential gene involved in checkpoint control and regulation of the MEC1 signalling pathway in *Saccharomyces cerevisiae*. *EMBO J.* **19**, 5801–5812 (2000).
42. Xu, X. *et al.* Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform-deficient cells. *Mol. Cell* **3**, 389–395 (1999).
43. Wang, Y. *et al.* BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev.* **14**, 927–939 (2000).
44. Gong, J. G. *et al.* The tyrosine kinase c-Abl regulates p73 in apoptotic response to cisplatin-induced DNA damage. *Nature* **399**, 806–809 (1999).
45. Kastan, M. B. *et al.* A mammalian cell cycle checkpoint pathway utilizing p53 and GADD45 is defective in ataxia-telangiectasia. *Cell* **71**, 587–597 (1992).
46. Paules, R. S. *et al.* Defective G2 checkpoint function in cells from individuals with familial cancer syndromes. *Cancer Res.* **55**, 1763–1773 (1995).
47. Cliby, W. A. *et al.* Overexpression of a kinase-inactive ATR protein causes sensitivity to DNA-damaging agents and defects in cell cycle checkpoints. *EMBO J.* **17**, 159–169 (1998).
48. Wright, J. A. *et al.* Protein kinase mutants of human ATR increase sensitivity to UV and ionizing radiation and abrogate cell cycle checkpoint control. *Proc. Natl. Acad. Sci. USA* **95**, 7445–7450 (1998).
49. Kim, S. T., Lim, D. S., Canman, C. E. & Kastan, M. B. Substrate specificities and identification of putative substrates of ATM kinase family members. *J. Biol. Chem.* **274**, 37538–37543 (1999).
50. Canman, C. E. *et al.* Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* **281**, 1677–1679 (1998).
51. Banin, S. *et al.* Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science* **281**, 1674–1677 (1998).
52. Tibbetts, R. S. *et al.* A role for ATR in the DNA damage-induced phosphorylation of p53. *Genes Dev.* **13**, 152–157 (1999).
53. Tibbetts, R. S. *et al.* Functional interactions between BRCA1 and the checkpoint kinase ATR during genotoxic stress. *Genes Dev.* (in press).
54. Barlow, C. *et al.* Atm-deficient mice: a paradigm of ataxia telangiectasia. *Cell* **86**, 159–171 (1996).
55. Elson, A. *et al.* Pleiotropic defects in ataxia-telangiectasia protein-deficient mice. *Proc. Natl. Acad. Sci. USA* **93**, 13084–13089 (1996).
56. Desany, B. A., Alcasabas, A. A., Bachant, J. B. & Elledge, S. J. Recovery from DNA replication stress is the essential function of the S-phase checkpoint pathway. *Genes Dev.* **12**, 2956–2970 (1998).
57. Khanna, K. K. *et al.* ATM associates with and phosphorylates p53: mapping the region of interaction. *Nature Genet.* **20**, 398–400 (1998).
58. Chan, D. W. *et al.* Purification and characterization of ATM from human placenta. A manganese-dependent, wortmannin-sensitive serine/threonine protein kinase. *J. Biol. Chem.* **275**, 7803–7810 (2000).
59. Smith, G. C. *et al.* Purification and DNA binding properties of the ataxia-telangiectasia gene product ATR. *Proc. Natl. Acad. Sci. USA* **96**, 11134–11139 (1999).
60. Lakin, N. D., Hann, B. C. & Jackson, S. P. The ataxia-telangiectasia related protein ATR mediates DNA-dependent phosphorylation of p53. *Oncogene* **18**, 3989–3995 (1999).
61. Hall-Jackson, C. A., Cross, D. A., Morrice, N. & Smythe, C. ATR is a caffeine-sensitive, DNA-activated protein kinase with a substrate specificity distinct from DNA-PK. *Oncogene* **18**, 6707–6713 (1999).
62. Matsuoka, S., Huang, M. & Elledge, S. J. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science* **282**, 1893–1897 (1998).
63. Blasina, A. *et al.* A human homologue of the checkpoint kinase Cds1 directly inhibits Cdc25 phosphatase. *Curr. Biol.* **9**, 1–10 (1999).
64. Brown, A. L. *et al.* A human Cds1-related kinase that functions downstream of ATM protein in the cellular response to DNA damage. *Proc. Natl. Acad. Sci. USA* **96**, 3745–3750 (1999).
65. Chaturvedi, P. *et al.* Mammalian Chk2 is a downstream effector of the ATM-dependent DNA damage checkpoint pathway. *Oncogene* **18**, 4047–4054 (1999).
66. Zhou, B.-B. S. *et al.* Caffeine abolishes the mammalian G2/M DNA damage checkpoint by inhibiting ataxia-telangiectasia-mutated kinase activity. *J. Biol. Chem.* **275**, 10342–10348 (2000).
67. Matsuoka, S. *et al.* ATM phosphorylates Chk2 in vivo and in vitro. *Proc. Natl. Acad. Sci. USA* **97**, 10389–10394 (2000).
68. Melchionna, R., Chen, X., Blasina, A. & McGowan, C. H. Threonine 68 is required for radiation-induced phosphorylation and activation of Cds1. *Nature Cell Biol.* **2**, 762–765 (2000).
69. Chehab, N. H., Malikzay, A., Appel, M. & Halazonetis, T. D. Chk2/hCds1 functions as a DNA damage checkpoint in G(1) by stabilizing p53. *Genes Dev.* **14**, 278–288 (2000).
70. Shieh, S. Y., Ahn, J., Tamai, K., Taya, Y. & Prives, C. The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites. *Genes Dev.* **14**, 289–300 (2000).
71. Chehab, N. H., Malikzay, A., Stavridi, E. S. & Halazonetis, T. D. Phosphorylation of Ser-20 mediates stabilization of human p53 in response to DNA damage. *Proc. Natl. Acad. Sci. USA* **96**, 13777–13782 (1999).
72. Ashcroft, M., Kubbutat, M. H. & Vousden, K. H. Regulation of p53 function and stability by phosphorylation. *Mol. Cell. Biol.* **19**, 1751–1758 (1999).
73. Khosravi, R. *et al.* Rapid ATM-dependent phosphorylation of MDM2 precedes p53 accumulation in response to DNA damage. *Proc. Natl. Acad. Sci. USA* **96**, 14973–14977 (1999).
74. Bell, D. W. *et al.* Heterozygous germ line *hCHK2* mutations in Li-Fraumeni syndrome. *Science* **286**, 2528–2531 (1999).
75. Lee, J. S., Collins, K. M., Brown, A. L., Lee, C. H. & Chung, J. H. hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response. *Nature* **404**, 201–204 (2000).
76. Sanchez, Y. *et al.* Conservation of the Chk1 checkpoint pathway in mammals: linkage of DNA damage to Cdk regulation through Cdc25. *Science* **277**, 1497–1501 (1997).
77. Walworth, N. C. & Bernards, R. rad-dependent response of the chk1-encoded protein kinase at the DNA damage checkpoint. *Science* **271**, 353–356 (1996).
78. Nurse, P. Checkpoint pathways come of age. *Cell* **91**, 865–867 (1997).
79. Morgan, D. O. Principles of CDK regulation. *Nature* **374**, 131–134 (1995).

80. Peng, C. Y. *et al.* Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216. *Science* **277**, 1501–1505 (1997).
81. Dalal, S. N., Schweitzer, C. M., Gan, J. & DeCaprio, J. A. Cytoplasmic localization of human cdc25C during interphase requires an intact 14-3-3 binding site. *Mol. Cell. Biol.* **19**, 4465–4479 (1999).
82. Yang, J., Winkler, K., Yoshida, M. & Kornbluth, S. Maintenance of G2 arrest in the *Xenopus* oocyte: a role for 14-3-3-mediated inhibition of Cdc25 nuclear import. *EMBO J.* **18**, 2174–2183 (1999).
83. Bunz, F. *et al.* Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science* **282**, 1497–1501 (1998).
84. Chan, T. A., Hermeking, H., Lengauer, C., Kinzler, K. W. & Vogelstein, B. 14-3-3 Sigma is required to prevent mitotic catastrophe after DNA damage. *Nature* **401**, 616–620 (1999).
85. Jelinsky, S. A. & Samson, L. D. Global response of *Saccharomyces cerevisiae* to an alkylating agent. *Proc. Natl Acad. Sci. USA* **96**, 1486–1491 (1999).
86. Hwang, B. J., Ford, J. M., Hanawalt, P. C. & Chu, G. Expression of the p48 xeroderma pigmentosum gene is p53-dependent and is involved in global genomic repair. *Proc. Natl Acad. Sci. USA* **96**, 424–428 (1999).
87. Tanaka, H. *et al.* A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage. *Nature* **404**, 42–49 (2000).
88. Carney, J. P. *et al.* The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell* **93**, 477–486 (1998).
89. Stewart, G. S. *et al.* The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. *Cell* **99**, 577–587 (1999).
90. Jeggo, P. A., Carr, A. M. & Lehmann, A. R. Splitting the ATM: distinct repair and checkpoint defects in ataxia-telangiectasia. *Trends Genet.* **14**, 312–316 (1998).
91. Moynahan, M. E., Chiu, J. W., Koller, B. H. & Jasin, M. Brca1 controls homology-directed DNA repair. *Mol. Cell* **4**, 511–518 (1999).
92. Bashkurov, V. I., King, J. S., Bashkurova, E. V., Schmuckli-Maurer, J. & Heyer, W. D. DNA repair protein Rad55 is a terminal substrate of the DNA damage checkpoints. *Mol. Cell. Biol.* **20**, 4393–4404 (2000).
93. Brush, G. S., Morrow, D. M., Hieter, P. & Kelly, T. J. The ATM homologue MEC1 is required for phosphorylation of replication protein A in yeast. *Proc. Natl Acad. Sci. USA* **93**, 15075–15080 (1996).
94. Lydall, D. & Weinert, T. Yeast checkpoint genes in DNA damage processing: implications for repair and arrest. *Science* **270**, 1488–1491 (1995).
95. Bochar, D. A. *et al.* BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. *Cell* **102**, 257–265 (2000).
96. Gasser, S. M. A sense of the end. *Science* **288**, 1377–1379 (2000).

Acknowledgements

We thank A. Carr, D. Cortez, M. Mattern, M. Huang, L. Zou, M. Tetzlaff and T. Weinert for comments on the manuscript.

Correspondence should be addressed to S. J. E. (e-mail: selledge@bcm.tmc.edu).

Optical gain in silicon nanocrystals

L. Pavesi*, L. Dal Negro*, C. Mazzoleni*, G. Franzò† & F. Priolo†

* INFN & Dipartimento di Fisica, Università di Trento, Via Sommarive 14, 38050 Povo, Italy

† INFN & Dipartimento di Fisica e Astronomia, Università di Catania, Corso Italia 57, 95129 Catania, Italy

Adding optical functionality to a silicon microelectronic chip is one of the most challenging problems of materials research. Silicon is an indirect-bandgap semiconductor and so is an inefficient emitter of light. For this reason, integration of optically functional elements with silicon microelectronic circuitry has largely been achieved through the use of direct-bandgap compound semiconductors. For optoelectronic applications, the key device is the light source—a laser. Compound semiconductor lasers exploit low-dimensional electronic systems, such as quantum wells and quantum dots, as the active optical amplifying medium. Here we demonstrate that light amplification is possible using silicon itself, in the form of quantum dots dispersed in a silicon dioxide matrix. Net optical gain is seen in both waveguide and transmission configurations, with the material gain being of the same order as that of direct-bandgap quantum dots. We explain the observations using a model based on population inversion of radiative states associated with the Si/SiO₂ interface. These findings open a route to the fabrication of a silicon laser.

Silicon, the mainstay semiconductor in microelectronic circuitry, has been considered unsuitable for optoelectronic applications owing to its indirect electronic bandgap, which limits its efficiency as a light emitter. Recently, room-temperature light emission from silicon has been shown to be possible when the silicon is in the form of a low-dimensional system^{1–5} or when selected active impurities (such as erbium⁶) and/or new phases (such as iron disilicide⁷) are inserted into the silicon lattice. All manner of low-dimensional silicon systems—such as porous silicon^{1,2,5}, silicon nanocrystals³, silicon/insulator superlattices⁴, silicon nano-pillars⁸—are being actively investigated as a means of improving the light-emission properties of silicon. The physical mechanism underlying high external quantum efficiencies for photoluminescence in low-dimensional silicon is mainly that of the quantum confinement of excitons in a nanometre-scale crystalline structure⁹, although the silicon/dielectric interface is also thought to play an active role in both the passivation of non-radiative states and the formation of radiative states¹⁰. Such work has led to many claims of a future role for silicon in photonic applications^{9,11–13}, yet a silicon laser has remained unlikely¹⁴.

To produce a silicon-based laser, we should demonstrate its light amplification or stimulated emission¹⁵. But light amplification in silicon is difficult because (1) it has efficient free carrier absorption, which reduces the net gain available for laser action¹⁴; (2) there is

significant Auger saturation of the luminescence intensity at high power⁹; and (3) there is significant size-dependence of the radiative energies in Si nanostructures, which yields large inhomogeneous broadening and significant optical losses in the system¹⁶. Here we report measurements of stimulated emission and light amplification in Si nanostructures and demonstrate optical gain in a single pass configuration. Population inversion is realized between the fundamental and a radiative state associated with the nanocrystal–oxide interface¹⁰. These findings could lead the way to a silicon-based laser.

Silicon nanocrystals

Low-dimensional silicon nanocrystals have been produced by negative ion implantation (80 keV; 1×10^{17} Si ions cm⁻²) into ultra-pure quartz substrates or into thermally grown silicon dioxide layers on Si substrates, followed by high-temperature thermal annealing (1,100 °C for 1 h). Quartz wafers (hereafter referred to

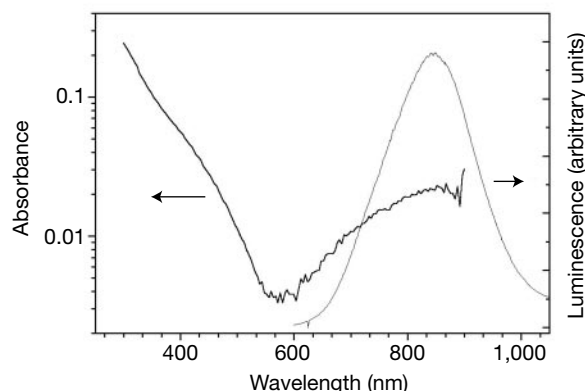


Figure 1 Room temperature absorbance and luminescence of Si nanocrystals embedded in a quartz matrix. The experimental set-up limited the absorbance measurement range. The absorbance of the quartz wafer was subtracted from the measured spectra. The 488 nm line of an Ar laser excited the luminescence.

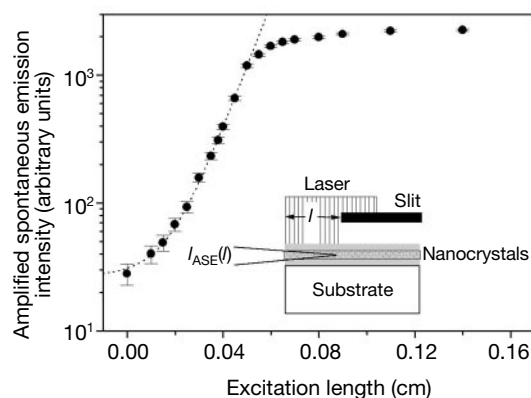


Figure 2 Amplified spontaneous emission intensity (ASE, disks) versus excitation stripe length (l) of Si nanocrystals embedded in a quartz matrix. Recording wavelength, 800 nm. A fit to the data with equation (1) is shown as the dashed line. The inset shows the experimental method. A cylindrical lens was used to focus the laser beam on the sample surface on a stripe 10- μ m wide and of variable length. Only the central part of the laterally unfocused laser spot was used to excite the sample. Measurements show that within these experimental conditions the laser power density on the sample surface is constant and independent of l . An optical 40 $\times$ objective imaged the sample edge on a 40- μ m pinhole so that only the light coming from the near sample surface region was collected. The use of a pulsed laser avoided any thermal heating of the sample. The excitation conditions were 1 kW cm⁻² mean power at a wavelength of 390 nm.

as 'sample A') were used for optical transmission experiments, and silicon wafers (hereafter referred to as 'sample B') to demonstrate microelectronics compatibility. Transmission electron microscopy of these samples showed silicon nanocrystals embedded within the oxide matrix. They were formed in a region centred at a depth of 110 nm from the sample surface and extending for a thickness of 100 nm; they were ~ 3 nm in diameter, with a concentration of $2 \times 10^{19} \text{ cm}^{-3}$. If we consider the Maxwell–Garnett approximation, we can estimate an effective refractive index of 1.89 for the nanocrystal region (see Fig. 3 in the Supplementary Information). For nanocrystals produced by plasma-enhanced chemical vapour deposition (PE-CVD)¹⁷, an effective refractive index of 1.71 was measured by ellipsometry. We note that in sample A, this causes the formation of a planar waveguide with an optical filling factor of about 9.7% when a refractive index of 1.89 is considered, or of 1.17% when a refractive index of 1.71 is assumed (see Fig. 3 in the Supplementary Information).

Absorbance and luminescence spectra at room temperature for sample A are shown in Fig. 1. A single wide emission band peaked at 800 nm, characteristic of the radiative recombination of carriers in Si nanocrystals, is observed. Absorbance measurements revealed a band in the near-infrared and a rising absorption edge at shorter wavelengths. The rising edge is due to absorption in the quantum confined states of the nanocrystals¹⁸, whereas the peculiar feature of the near-infrared absorption band is caused by a Si=O interface state^{10,19,20}. As predicted by theory^{10,21} and inferred from experiment, the interface state is formed at the interface between the Si nanocrystals and the SiO₂ matrix. The microscopic nature of these interface states is still under debate^{10,21}. Very good quality SiO₂ and Si nanocrystals are needed to observe this interface state, which in other Si-based systems is hindered by interfaces with defects or the low quality of the oxide. We note the spectral coincidence of the emission band and the interface state absorption band, suggesting that radiative emission in Si nanocrystals occurs through a radiative state associated with the nanocrystal–oxide interface. Time-resolved luminescence, under picosecond excitation, on our nanocrystals showed a very fast rise time, within our experimental sensitivity (some nanoseconds)²². The decay time of the luminescence was in the microsecond range; it is dependent on the emission energy²³.

Light amplification

To measure light amplification we used the variable strip length method (see the inset of Fig. 2)²⁴. The sample is optically excited by a

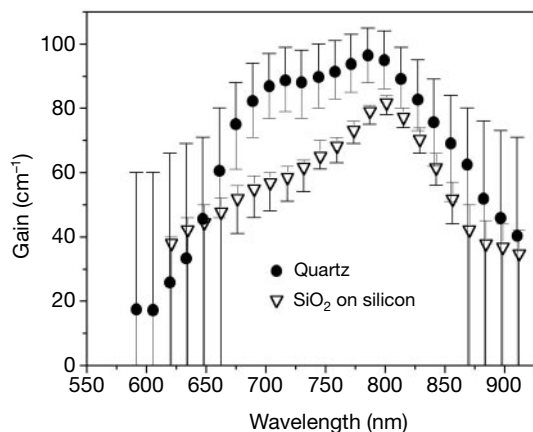


Figure 3 Spectral dependence of the net modal gain. Sample A, circles; sample B, triangles. The experimental conditions were as in Fig. 2. The large error bars result from both the low signal-to-noise ratio of the streak camera detection for low intensity signals and from the numerical procedure used for obtaining the modal gain from the ASE data^{30,31}.

doubled Ti:sapphire laser beam ($\lambda = 390$ nm, 2-ps pulse width, 82-MHz repetition rate) in a stripe-like geometry with variable length (l). The amplified spontaneous emission intensity I_{ASE} that is emitted from the sample edge (observation angle $\phi = 0$) is measured as a function of l . From a fit of the resulting curve, the optical gain g can be deduced at every wavelength. By assuming a one-dimensional amplifier model, I_{ASE} can be related to g by^{15,24}

$$I_{\text{ASE}}(l) \propto \frac{I_{\text{SPONT}} \times l}{g - \alpha} (e^{(g - \alpha)l} - 1) \quad (1)$$

where I_{SPONT} is the spontaneous emission intensity per unit length and α an overall loss coefficient. The gain measured in this way is the modal gain¹⁵. Figure 2 shows I_{ASE} versus l in Si nanocrystals measured at a wavelength of 800 nm. For small values of l (< 0.05 cm), an exponential increase of I_{ASE} is observed that indicates the occurrence of amplified spontaneous emission. A fit with equation (1) yields the net modal gain $g - \alpha = 100 \pm 10 \text{ cm}^{-1}$. For large values of l (> 0.05 cm), I_{ASE} saturates as expected for any finite power supply amplification mechanism. At

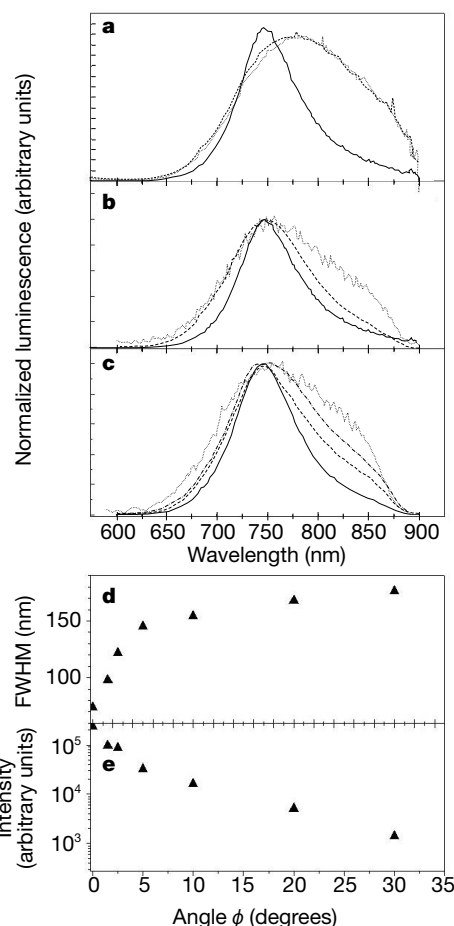


Figure 4 Amplified spontaneous emission spectra of sample A for different measurement conditions. **a**, Amplified spontaneous emission (ASE) spectra for a constant excitation length $l = 2,000 \mu\text{m}$ and various power densities P : continuous line 2.3 kW cm^{-2} , dashed line 1 kW cm^{-2} , dotted line 170 W cm^{-2} . **b**, ASE spectra for constant $P = 2.3 \text{ kW cm}^{-2}$ and various l values: $l = 2,000 \mu\text{m}$ continuous line, $l = 650 \mu\text{m}$ dashed line, $l = 200 \mu\text{m}$ dotted line. **c**, ASE spectra for constant $P = 1 \text{ kW cm}^{-2}$ and $l = 3,000 \mu\text{m}$ and various observation angles ϕ . ϕ is defined with respect to the optical axis of the one-dimensional amplifier. Continuous line $\phi = 0^\circ$, dashed line $\phi = 1.5^\circ$, dash-dotted line $\phi = 2.5^\circ$, dotted line $\phi = 20^\circ$. The low energy cut-off of the photomultiplier used to record the data deforms the spectra for wavelength longer than 880 nm. **d**, Full-width at half-maximum (FWHM) of the ASE emission as a function of the observation angle ϕ . **e**, Peak intensity of the ASE signal as a function of ϕ .

low power density, we measured absorption; when the pump power was increased, the peak net modal gain increased and then saturated at values of about 100 cm^{-1} for power densities of about 5 kW cm^{-2} .

By measuring the amplified signal for various wavelengths we obtained the gain spectrum for both samples A and B (Fig. 3; see also Supplementary Information). A wide spectral band is observed which spectrally overlaps the wavelength range of the luminescence, demonstrating that amplification is produced by the radiative state associated with the nanocrystal–oxide interface. We noticed that both samples yielded similar shapes and values for the gain curve. A confirmation of these findings was the observation of a strong emission lineshape narrowing (Fig. 4), either when the pump power density P is increased with fixed excitation length l (Fig. 4a), or when the excitation length l is increased with a fixed P (Fig. 4b). When l and P are fixed and the observation angle ϕ is changed (Fig. 4c–e), a significant intensity decrease and a broadening of the amplified emission spectrum occur as soon as there is deviation from the strict one-dimensional amplifier configuration, that is, when $\phi > 0^\circ$. These observations support also the waveguide formation in our samples.

The most direct evidence of light amplification from our systems was provided by pump and probe transmission measurements. An intense laser beam (pump) at 390 nm excites the sample in order to reach the population inversion needed for amplification, while a weak probe signal at $\sim 800 \text{ nm}$ passes through the active layer of thickness d . In the presence (absence) of the pump beam the probe beam is amplified (absorbed). In Fig. 5 we show the results. The probe signal is clearly amplified when passing through the excited nanocrystals. To our knowledge, this is the first evidence of light amplification in transmission, usually named single-pass gain, in Si-based systems. We deduced the net material gain values by using the formula given in Fig. 5; they are high enough to compare with those of self-assembled quantum dots made of III–V semiconductors^{25,26}: $10,000 \pm 3,000 \text{ cm}^{-1}$. This value has a very large error bar because of the geometry of the active nanocrystal layer and the losses in the quartz substrate. No change in probe intensity in the presence/absence of the pump beam was observed when the probe beam passed through pure quartz (without nanocrystals). Moreover, by decreasing the pump intensity (Fig. 5, right panel) we measured even absorption of the probe beam (population inversion is no

longer reached in nanocrystals). By changing the probe wavelength the net material gain decreased and eventually disappeared (amplification is lost when the probe energy is no longer in resonance with the transition for which population inversion is achieved) with an overall spectral dependence similar to that shown in Fig. 3 (see also Supplementary Information).

Gain cross-section per nanocrystal

By using the formalism of ref. 27 and the measured probe beam transmission under inversion conditions, we estimated a maximum-gain cross-section per nanocrystal $\gamma_T \approx 5 \times 10^{-16} \text{ cm}^2$. It is interesting to compare this gain cross-section per nanocrystal with the photon absorption cross-section per nanocrystal (σ). We have hence directly measured the absorption cross-sections of ion-implanted Si nanocrystals by studying the rise time of the photoluminescence intensity as a function of pump power in a fashion similar to that recently reported for porous Si²⁸.

The photoluminescence intensity is given by $I \propto N^*/\tau_R$, where N^* is the concentration of excited nanocrystals and τ_R the radiative lifetime. The rate equation for nanocrystal excitation will be:

$$\frac{dN^*}{dt} = \sigma J(N - N^*) - \frac{N^*}{\tau} \quad (2)$$

where J is the photon flux, N is the concentration of nanocrystals and τ is the decay time, taking into account both radiative and non-radiative processes. If a continuous wave (CW) pumping laser is turned on at $t = 0$, N^* will change according to equation (2) and the photoluminescence intensity will increase according to the following law:

$$I(t) = I_0 \left\{ 1 - \exp \left[- \left(\sigma J + \frac{1}{\tau} \right) t \right] \right\} = I_0 \left\{ 1 - \exp \left[- \left(\frac{t}{\tau_{\text{on}}} \right) \right] \right\} \quad (3)$$

A measure of the photoluminescence rise time as a function of J will therefore give direct information on the absorption cross-section. The inset to Fig. 6 shows $I(t)$ at 850 nm for Si nanocrystals pumped at 488 nm for different pump powers P . As predicted by equation (3), the photoluminescence rise time becomes shorter and shorter

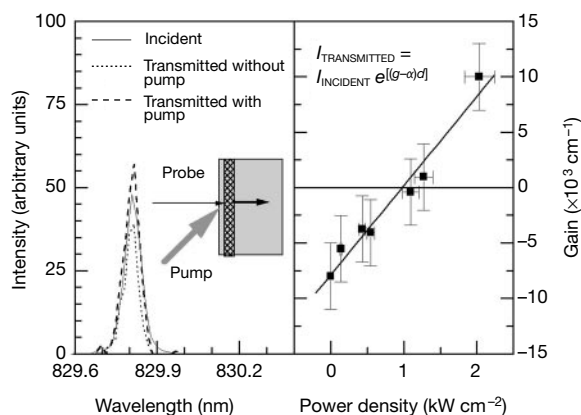


Figure 5 Gain measurements. Left panel, spectrum of the transmitted probe beam measured in presence (dashed line) or in absence (dotted line) of the pump beam. The spectrum of the transmitted probe beam in absence of the absorbing/amplifying nanocrystal medium is also reported (continuous line) and named the incident beam. The inset shows the principle of the experiment. The formula used to deduce the material gain is also reported, d being the thickness of the active region. The probe beam was provided by a Kr lamp, which was imaged to a spot size of about 0.01 mm^2 on the sample surface. The pump beam had a mean power of about 2 kW cm^{-2} and a wavelength of 390 nm. Right panel, dependence of the material gain value on the pump power density.

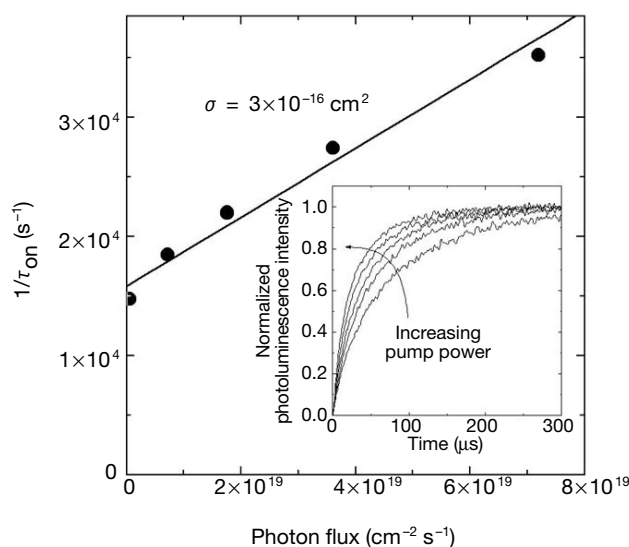


Figure 6 Reciprocal of the rise time τ_{on} as a function of the pump laser photon flux as obtained from a fit to the time resolved photoluminescence data shown in the inset. The slope gives the photon absorption cross-section σ . Inset: time resolved photoluminescence intensity at 850 nm switching on the 488 nm line of a CW Ar pumping laser at $t = 0$; data are taken at room temperature, at different pump powers (in the 0.8–80 mW range) and are normalized to the maximum intensity.

as P is increased. By fitting these curves with equation (3) we obtain the values of the rise time, τ_{on} , at the different P values. The reciprocal of τ_{on} is reported in Fig. 6 as a function of J . The data follow a straight line with a slope $\sigma \approx 3 \times 10^{-16} \text{ cm}^2$. The intercept of the fitted straight line with the vertical axis gives the lifetime of the Si nanocrystals in the system at the measured wavelength. The obtained value (70 μs) is in agreement with decay time measurements at 850 nm on the same sample. In this way we have been able to obtain a direct measurement of the photon absorption cross-section of the nanocrystals. Although this measurement is performed at an excitation wavelength of 488 nm, it should reflect the property of the 800 nm state because absorbance at these two wavelengths is identical (see Fig. 1). We note that, as theoretically predicted¹⁵, the measured absorption cross-section σ is of the same order of magnitude as the gain cross-section γ_{T} . The same agreement is found when we compare the net material gain to the absorption coefficient deduced by the absorbance data of Fig. 1.

Another important issue concerns the comparison of the gain cross-sections that are derived from the modal and the material gain. It was shown in ref. 27 that in the variable stripe-length geometry, the gain cross-section per nanocrystal (γ_{ASE}) can be derived by using

$$\gamma_{\text{ASE}} = \frac{g}{(f_c - f_v)N\Gamma} \quad (4)$$

where Γ is the optical filling factor of the amplified mode. By assuming a complete population inversion $f_c - f_v = 1$, an optical filling factor of 0.097 and the measured net modal gain $g \approx 100 \text{ cm}^{-1}$, one finds $\gamma_{\text{ASE}} \approx 5 \times 10^{-17} \text{ cm}^2$. This is a lower limit to γ_{ASE} . If we consider an incomplete inversion or a weaker confining waveguide (that is, no step index profile in the waveguide and/or a lower effective refractive index for the nanocrystal), γ_{ASE} will significantly increase. For example, by using the Γ value computed for the refractive index measured in PE-CVD nanocrystals¹⁷, we obtain $\gamma_{\text{ASE}} \approx 3 \times 10^{-16} \text{ cm}^2$. The difference between the two estimated values of γ_{ASE} may be an indication of the quantitative uncertainty on the gain cross-section values determined by the gain measurements.

A comparison of the gain cross-sections per nanocrystal derived by transmission (γ_{T}) and by the variable strip length method (γ_{ASE}) shows that the two values are in reasonable quantitative agreement. Indeed, the determined value of γ_{ASE} is only a lower limit and there

are large error bars on the pump and probe determined gain coefficients.

Origin of gain

The wide spectral gain of inhomogeneous nature is energetically matching both the luminescence emission and the 800-nm interface state absorption band. For these reasons, a three-level model is proposed to explain the observed gain (Fig. 7): two levels correspond to the lowest unoccupied molecular orbital (LUMO), or the bottom of the conduction band, and to the highest occupied molecular orbital (HOMO), or the top of the valence band of the nanocrystal, respectively. The third level is due to the radiative interface state observed in absorption and responsible for the luminescence emission band at 800 nm. Optical excitation populates the LUMO, emptying the HOMO. Electrons from the LUMO relax very rapidly to the interface state. Electrons in the interface state have long lifetimes. Indeed, the absorption band at 800 nm, the Stokes shift between absorption and luminescence, the fast rise and the slow decay times of the 800-nm luminescence under picosecond excitation, and the efficient luminescence emission of the 800-nm luminescence all support this energy model. Within the model, the rate of depopulation of the initial state is much faster than its filling rate via a carrier recombination mediated by the interface state. Population inversion between the HOMO and the radiative state associated with the nanocrystal–oxide interface is thus possible. This model also explains why losses due to free carrier absorption that usually exceed the gain by stimulated emission in other Si based systems²⁹, or those due to Auger recombinations²², are not effective here. In addition, model calculations¹⁰ show that the size dependence of the radiative interface state energy is smaller than that of conduction-to-valence band transitions, relaxing issues related to the broad distribution of sizes.

Using the measured absorption cross-section σ per nanocrystal, we estimate that under our peak excitation condition of about 10^{22} photons $\text{cm}^{-2} \text{ s}^{-1}$, more than 100 electron–hole pairs per nanocrystal are generated. As we have nanocrystals with about 500 Si atoms, of which about 35% are surface atoms, we have about 150 interface states available per Si nanocrystal when we assume that each surface Si atom is bound to an O atom. In Fig. 5, we show that to have optical gain the excitation level should be high enough to invert most of these states.

In Table 1 we report a compilation of data on the gain cross-section per quantum dot for some III–V semiconductors. It can be noticed that the silicon nanocrystal values are about three orders of magnitude lower than the one typically found in InAs quantum dots. We argue that this is due to the indirect bandgap of Si and to the fact that the gain is due to radiative interface states. Despite this difference, the net material gain is of the same order of magnitude between Si nanocrystals and InAs quantum dot systems, owing to the much higher areal density of nanocrystals that is achievable with the ion-implantation method used in this work. We note that the gain cross-section per nanocrystal γ is inversely proportional to the radiative lifetime τ_{R} (ref. 15). By looking at the γ data in Table 1, we infer large differences in lifetimes. Indeed radiative lifetimes in the microsecond range are measured for Si nanocrystals, whereas for InAs quantum dot lifetimes in the nanosecond range were reported^{25,26}.

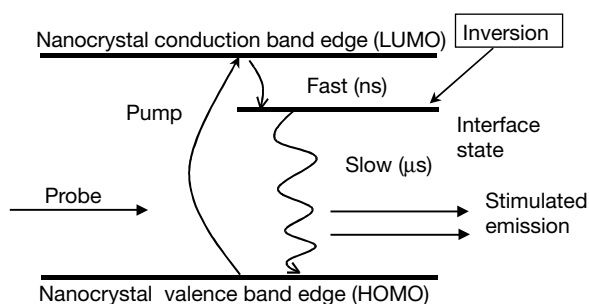


Figure 7 Schematic energy diagram for a nanocrystal showing how population inversion can be reached in this system.

Table 1 Gain cross-section per quantum dot or nanocrystal

Quantum dot material	Net model gain (cm^{-1})	Net material gain ($\times 10^4 \text{ cm}^{-1}$)	Areal dot density (cm^{-2})	Active layer thickness (nm)	Filling factor (10^{-4})	Gain cross-section per dot (10^{-16} cm^2)	Reference
InAs single layer quantum dot	8.2	9*	1×10^{11}	1.7	1.2	1,200	25
InAs 7 stacks quantum dot	70–85	1.5*	1×10^{11}	100	48	4,000	26
GaAs single layer quantum dot	13*		1×10^{10}			450*	27
Si nanocrystals	100	1	2×10^{14}	100	970	0.5–5	This work

* Calculated approximately.

Conclusions

Modal and net material optical gains have been observed unambiguously in Si nanocrystals. Quantitative estimates of gain cross-section per nanocrystal show that the measured values are orders of magnitude lower than those found in III–V semiconductor quantum dots. However, owing to the much higher stacking density of Si nanocrystals with respect to direct-bandgap quantum dots, similar values for the material gain are observed. These findings open a route towards the realization of a silicon-based laser. □

Received 8 June; accepted 24 October 2000.

- Canham, L. T. Silicon quantum wire array fabrication by electrochemical and chemical dissolution of wafers. *Appl. Phys. Lett.* **57**, 1045–1048 (1990).
- Cullis, A. G. & Canham, L. T. Visible light emission due to quantum size effects in highly porous crystalline silicon. *Nature* **353**, 335–338 (1991).
- Wilson, W. L., Szajowski, P. F. & Brus, L. E. Quantum confinement in size-selected surface-oxidised silicon nanocrystals. *Science* **262**, 1242–1244 (1993).
- Lu, Z. H., Lockwood, D. J. & Baribeau, J.-M. Quantum confinement and light emission in SiO₂/Si superlattices. *Nature* **378**, 258–260 (1995).
- Hirschman, K. D., Tsybeskov, L., Duttagupta, S. P. & Fauchet, P. M. Silicon-based light emitting devices integrated into microelectronic circuits. *Nature* **384**, 338–340 (1996).
- Franzò, G., Priolo, F., Coffa, S., Polman, A. & Carnera, A. Room temperature electroluminescence from Er doped crystalline silicon. *Appl. Phys. Lett.* **64**, 2235–2237 (1994).
- Leong, D., Harry, M., Reeson, K. J. & Homewood, K. P. A silicon/iron disilicide light-emitting diode operating at a wavelength of 1.5 µm. *Nature* **387**, 686–688 (1997).
- Nassiopoulos, A. G., Grigoropoulos, S. & Papadimitriou, D. Electroluminescent device based on silicon nanopillars. *Appl. Phys. Lett.* **69**, 2267–2269 (1996).
- Bisi, O., Ossicini, S. & Pavesi, L. Porous silicon: a quantum sponge structure for silicon based optoelectronics. *Surf. Sci. Rep.* **38**, 1–126 (2000).
- Wolkin, M. V., Jorne, J., Fauchet, P. M., Allan, G. & Delerue, C. Electronic states and luminescence in porous silicon quantum dots: the role of oxygen. *Phys. Rev. Lett.* **82**, 197–200 (1999).
- Miller, D. A. Silicon sees the light. *Nature* **378**, 238 (1995).
- Iyer, S. S. & Xie, Y.-H. Light emission from silicon. *Science* **260**, 40–46 (1993).
- Silicon based optoelectronics. *Mater. Res. Bull.* **23**(4), (1998).
- Fauchet, P. M. The integration of nanoscale porous silicon light emitters: materials science, properties and integration with electronic circuitry. *J. Lumin.* **80**, 53–64 (1999).
- Yariv, A. *Quantum Electronics* 2nd edn (Wiley & Sons, New York, 1974).
- Canham, L. T. in *Frontiers of Nano-Optoelectronic Systems* (eds Pavesi, L. & Buzaneva, E.) 85–98 (Kluwer Academic, Dordrecht, 2000).

- Iacona, F., Franzò, G. & Spinella, C. Correlation between luminescence and structural properties of Si nanocrystals. *J. Appl. Phys.* **87**, 1295–1303 (2000).
- Kovalev, D., Heckler, H., Polisski, G. & Koch, F. Optical properties of Si nanocrystals. *Phys. Status Solidi* **251**, 871–930 (1999).
- Kanemitsu, Y. & Okamoto, S. Quantum confinement and interface effects on photoluminescence from silicon single quantum wells. *Solid State Commun.* **103**, 573–576 (1997).
- Kanemitsu, Y. & Okamoto, S. Phonon structures and Stokes shift in resonantly excited luminescence of silicon nanocrystals. *Phys. Rev. B* **58**, 9652–9655 (1998).
- Degoli, E. & Ossicini, S. Quantum confined and interface states related visible luminescence in Si/SiO₂ superlattices. *Surf. Sci.* (in the press).
- Klimov, V. I., Schwarz, Ch., McBranch, D. W. & White, C. W. Initial carrier relaxation dynamics in ion-implanted Si nanocrystals: femtosecond transient absorption study. *Appl. Phys. Lett.* **73**, 2603–2605 (1998).
- Linnros, J., Galeckas, A., Lalic, N. & Grivickas, V. Time-resolved photoluminescence characterisation of nm-sized silicon nanocrystallites in SiO₂. *Thin Solid Films* **297**, 167–170 (1997).
- Shaklee, K. L., Nahaory, R. E. & Leheny, R. F. Optical gain in semiconductors. *J. Lumin.* **7**, 284–309 (1973).
- Kirstaedter, N. *et al.* Gain and differential gain of single layer InAs/GaAs quantum dot injection lasers. *Appl. Phys. Lett.* **69**, 1226–1228 (1996).
- Ling, C. *et al.* Dynamics of amplified spontaneous emission in InAs/GaAs quantum dots. *Appl. Phys. Lett.* **76**, 3507–3509 (2000).
- Blood, P. On the dimensionality of optical absorption, gain and recombination in quantum-confined structures. *IEEE J. Quantum Electron.* **36**, 354–362 (2000).
- Kovalev, D. *et al.* Optical absorption cross sections of Si nanocrystals. *Phys. Rev. B* **61**, 4485–4487 (2000).
- von Behren, J., Kostoulas, Y., Ucer, K. B. & Fauchet, P. M. The femtosecond optical response of porous, amorphous and crystalline silicon. *J. Non-Cryst. Solids* **198–200**, 957–960 (1996).
- Jordan, V. Gain measurements of semiconductor laser diodes: requirements for wavelength resolution and sensitivity to noise. *IEE Proc. Optoelectron.* **141**, 13–15 (1994).
- Hvam, J. M. Direct recording of optical gain spectra from ZnO. *J. Appl. Phys.* **49**, 3124–3126 (1978).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work has been supported by the National Institute for the Physics of the Matter (INFN) through the LUNA project and the advanced research project RAMSES.

Correspondence and request for materials should be addressed to L.P. (e-mail: pavesi@science.unitn.it).

Evolution of the Sun's large-scale magnetic field since the Maunder minimum

S. K. Solanki*, M. Schüssler* & M. Fligge†

* Max-Planck-Institut für Aeronomie, D-37191 Katlenburg-Lindau, Germany

† Institut für Astronomie, ETH-Zentrum, CH-8092 Zürich, Switzerland

The most striking feature of the Sun's magnetic field is its cyclic behaviour. The number of sunspots, which are dark regions of strong magnetic field on the Sun's surface, varies with a period of about 11 years. Superposed on this cycle are secular changes that occur on timescales of centuries and events like the Maunder minimum in the second half of the seventeenth century, when there were very few sunspots^{1,2}. A part of the Sun's magnetic field reaches out from the surface into interplanetary space, and it was recently discovered³ that the average strength of this interplanetary field has doubled in the past 100 years. There has hitherto been no clear explanation for this doubling. Here we present a model describing the long-term evolution of the Sun's large-scale magnetic field, which reproduces the doubling of the interplanetary field. The model indicates that there is a direct connection between the length of the sunspot cycle and the secular variations.

The magnetic field observed at the solar surface, a small part of which continues beyond the solar corona and permeates interplanetary space, is produced by a dynamo process operating in the solar interior. The field emerges at the surface in the form of bipolar regions embracing wide ranges of size and magnetic flux⁴; the larger of these are the active regions and the smallest are called ephemeral

regions. These and other concepts are illustrated in Fig. 1.

Individual elements of the magnetic field (magnetic flux tubes or bundles of field lines) are transported by flows at the solar surface and carry out a combination of systematic motion and random walk (diffusion), so that part of the magnetic flux of the active region is spread widely over the solar surface right up to the poles^{5,6}. The resulting large-scale pattern determines the global magnetic dipole moment and the strength of the field in interplanetary space^{7,8}. Although the total amount of flux appearing on the solar surface in ephemeral regions is much bigger⁹, the large-scale pattern of the field is mainly due to the large bipolar regions^{5,10}. In fact, the net contribution of the ephemeral regions to the axial dipole moment of the Sun is a factor of 6 smaller than that of the large active regions¹¹. Hence, for the simple approach taken here, it is sufficient to consider only the large active regions and follow the evolution of their magnetic flux. We are also restricted to the sunspot-bearing regions because the record of sunspot numbers is the only data set related to the solar magnetic field that reaches back to 1700. We note, however, that the influence of smaller bipolar regions is taken into account in a rough way; see equation (3) below.

In our model we consider the large-scale magnetic field (as measured, say, by the National Solar Observatory/Kitt Peak magnetograms¹²) to consist of two components: (1) (large) active regions and their early decay products (enhanced network, compare to Fig. 1); and (2) the part of the flux in the magnetic network that is organized into unipolar regions on a large enough scale for it to open into interplanetary space. As we are interested in the source of the interplanetary field we need not consider the closed part of the network flux.

Active region flux emerges rapidly (within days) and most of it is removed from the surface by cancellation at the border between opposite polarities within a few months¹³. We assume that a small part of the flux is transferred to the large-scale unipolar regions connected to the interplanetary field (which we call the open network) and resides there for a long time, of the order of years, in accordance with the observed long-term evolution of the large-scale field^{5,6,8}.

Owing to its short decay time, the (unsigned) magnetic flux in active regions, Φ_a , for our purpose can be assumed to be in quasi-steady balance between the relevant gain and loss terms: $E = T + \Phi_a/\tau_a$, where E is the flux emergence rate, $T = \Phi_a/\tau_t$ is

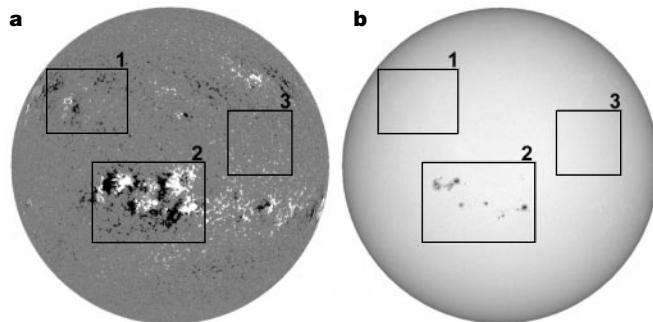


Figure 1 Distribution of the magnetic field and brightness on the solar surface. **a**, Magnetogram showing longitudinal magnetic field; **b**, filtergram showing brightness. Bright and dark features in the magnetograms represent magnetic flux of opposite polarity (that is, magnetic field directed out of and into the Sun, respectively); grey indicates the absence of net magnetic flux. The inserted frames 1–3 enclose features of particular interest. Frame 3 indicates a region of magnetic network (a 'salt and pepper' pattern in **a**, no dark sunspots in **b**). In frame 2, new magnetic flux has emerged from the solar interior to the surface and forms bipolar magnetic regions. These are identified by the large patches of field with opposite polarity and the sunspots visible in **b**. Only large bipolar regions form sunspots; they are called active regions. Their emergence rate varies strongly over the 11-year activity cycle of the Sun. Active regions begin to decay soon after the emergence of magnetic flux ceases. The sunspots break up and disappear within days or weeks, whereas the remainder of the active region often lives much longer in the form of the enhanced magnetic network (frame 1). Hence, the relative sunspot number, R , for which records exist back to 1700, can be used as a proxy for the flux emergence rate in studies of the long-term evolution of the large-scale solar magnetic field. Small bipolar regions without sunspots are referred to as ephemeral regions (data obtained by the Michelson-Doppler-Interferometer instrument²⁴ on board the Solar and Heliospheric Observatory, SOHO).

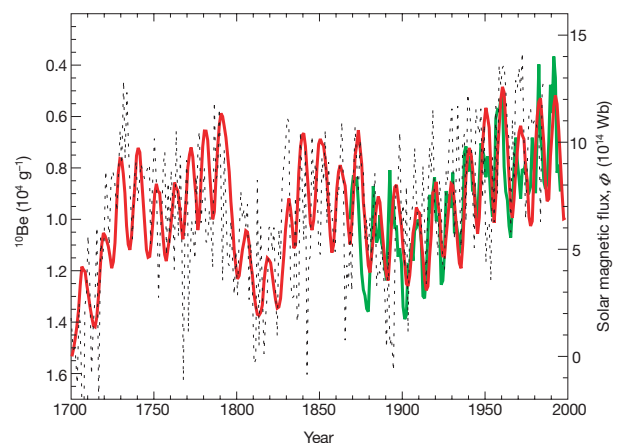


Figure 2 Evolution of the open magnetic flux at the solar surface since the end of the Maunder minimum in 1700. Data as predicted by our model are given by the red curve. For comparison, the flux of the interplanetary magnetic field³ reconstructed from the geomagnetic variations (green curve) and the ¹⁰Be concentration in ice cores² (dotted curve corresponding to the inverted scale on the left y-axis) are also plotted. The interplanetary flux values have been multiplied by a factor of 2 in order to obtain the total unsigned flux. The ¹⁰Be record has been plotted without any smoothing or filtering.

the rate at which flux is transferred from active regions to the open network, with τ_i being the corresponding timescale, and τ_a is the decay time due to cancellation of opposite-polarity magnetic flux in the active region. The resulting expression for T can be inserted as a source term into the equation for the (unsigned) open network flux, Φ_o , yielding

$$\frac{d\Phi_o}{dt} = \gamma E - \frac{\Phi_o}{\tau_o} \quad (1)$$

with $\gamma = \tau_a/(\tau_a + \tau_i)$; τ_o is the decay timescale of the open network flux. We fix the value of γ using the flux emergence rate for the maximum of solar cycle 21 given by ref. 13. Requiring equality of Φ_o and the total unsigned interplanetary flux of approximately 1.4×10^{15} Wb during the maximum³, we estimate γ by equating the two terms on the right-hand side of equation (1) for that value of Φ_o . This leaves τ_o as a free parameter of the model; the best fit to the reconstruction of ref. 3 is obtained with $\tau_o = 4$ yr (see below), which yields $\gamma \approx 0.015$.

The flux emergence rate in active regions, E , is expressed in terms of the relative sunspot number, R , as

$$E = c \left(1 + \frac{A_f}{A_s} \right) R \quad (2)$$

where c is a conversion factor which can be determined from empirical emergence rates. The factor $(1 + A_f/A_s)$ takes into account that only a part of the newly emerged flux is in the form of sunspots (covering the area A_s), whereas the major part is concentrated in small magnetic elements, ensembles of which are visible as bright faculae and cover the area A_f . The observed linear relationship between magnetic flux and the corresponding area covered by bright features outside sunspots¹³ supports our use of area as a proxy for magnetic flux. The ratio A_f/A_s in active regions varies over the solar cycle, being smaller at activity maximum than at minimum. Observations¹⁴ and reconstructions of short-term solar irradiance variations¹⁵ imply that:

$$\frac{A_f}{A_s}(R) = 21 + \frac{24.35}{R} - 0.061R \quad (3)$$

We note that A_f/A_s includes in a rough manner the contribution of small bipolar regions to flux emergence; through its nonlinear dependence on R it also takes into account to a certain extent that the number of small, relative to large, bipolar regions changes over the solar cycle. After substituting equations (2) and (3) into equation (1) the latter can be numerically integrated over any interval of time over which $R(t)$ is known.

If τ_o is sufficiently large, a secular variation of the open network flux is a natural consequence of our model. The cycle-averaged level of Φ_o depends not only on the total magnetic flux emerging during a cycle, Φ_{tot} , but also on the cycle length: shorter cycles lead to larger average open network flux (for constant Φ_{tot}). This can be simply illustrated by replacing the emergence term in equation (1) by a periodic function $a\Omega[1 + \cos(\Omega t)]$, where Ω is the cycle frequency. The amplitude factor, a , is multiplied by Ω in order to keep the total flux emerging per cycle independent of Ω . The resulting simplified version of equation (1) can be solved analytically. The open magnetic flux averaged over a cycle period asymptotically reaches the value:

$$\langle \Phi_o \rangle = a\Omega\tau_o \quad (4)$$

Hence, in addition to the cycle amplitude, the time-averaged open flux depends equally on the cycle length, with shorter cycles (larger Ω) yielding larger values of $\langle \Phi_o \rangle$. This effect is amplified by the anticorrelation between the length and the amplitude of the solar cycle^{16,17}.

In Fig. 2 we show Φ_o as reconstructed from the numerical solution of equation (1) since 1700, with $\tau_o = 4$ years (red curve). The year 1700 corresponds roughly to the end of the Maunder

minimum. At that time very few sunspots were present on the solar surface¹, so there was little flux emerging in active regions and sufficient time for the large-scale unipolar flux previously present on the solar surface to decay. We have therefore set Φ_o to zero at 1700. We compare our result with the reconstruction of the interplanetary magnetic field of ref. 3 (green curve). The value of the decay time was adjusted to $\tau_o = 4$ yr in order to make the model match the relative amplitudes of the cyclic to the secular change exhibited by the interplanetary magnetic field. The agreement between the two quantities is gratifying, given the simplicity of our model and the assumptions entering the reconstruction.

The dotted curve in Fig. 2 represents the ^{10}Be concentration deduced from an ice core drilled at Dye 3, Greenland². ^{10}Be is formed as a spallation product in the upper atmosphere of the Earth when cosmic rays hit nitrogen and oxygen nuclei. The ^{10}Be concentration is thus another measure of the solar magnetic activity, because a stronger, more complex solar magnetic field more efficiently shields the Earth from cosmic rays. Again, the agreement with our model is pleasing, and somewhat surprising in view of the unknown influence of past climate change on the ^{10}Be concentration. Not only are the secular behaviour of Φ_o and the ^{10}Be concentration almost identical, but also the ratio of cyclic to secular variations of Φ_o and the ^{10}Be data are compatible. This comparison also supports our choice of $\Phi_o = 0$ at 1700.

The resulting decay time of $\tau_o = 4$ yr allows for a significant redistribution of the open network flux by convection, differential rotation and meridional circulation during its lifetime^{7,8}. This is in accordance with the accumulation of magnetic flux at high heliolatitudes during solar minimum as inferred from Ulysses measurements¹⁸, which indicates that the residence time of magnetic flux must be sufficiently large for the meridional flow to concentrate the field at the poles. The polar fields dominate the interplanetary magnetic flux during solar activity minimum.

Our model yields a physical basis for understanding the secular trend of the interplanetary magnetic field detected in ref. 3. The decisive property of the model permitting the successful reproduction of the data is the long residence time of open magnetic flux at the solar surface. Considering only the instantaneous level of activity (say, the sunspot number) provides a very poor fit to the interplanetary field and the ^{10}Be data. Whereas the ^{10}Be data are dominated by a secularly changing background, the sunspot number always drops almost to zero at each activity minimum. An overlap between consecutive 'extended solar cycles'^{19,20} could, in principle, introduce a secular trend by raising the minima between successive strong cycles; however, the small bipolar regions (ephemeral regions) that are mainly responsible for this overlap contribute very little to the large-scale patterns of the solar magnetic field^{5,10} and the global dipole moment¹¹. At most, this effect could lead to some decrease of the value of τ_o .

The correspondence between our model and the ^{10}Be record provides compelling evidence that the latter indeed tracks the interplanetary field² and thus gives a sound basis for taking the ^{10}Be record as a proxy of solar activity in the past. The model also introduces a mechanism for linking the solar cycle length²¹ with secular variation of the interplanetary flux. Our reconstruction of the open solar magnetic flux provides the major parameter needed to reconstruct the secular variation of the cosmic ray flux impinging on the terrestrial atmosphere. It has been suggested that cosmic rays affect the total cloud cover of the Earth and thus drive the terrestrial climate^{22,23}. The work presented here will allow further examination of this proposed link. $\square$

Received 15 August; accepted 3 October 2000.

- Ribes, J. C. & Nesme-Ribes, E. The solar sunspot cycle in the Maunder minimum AD1645 to AD1715. *Astron. Astrophys.* **276**, 549–563 (1993).
- Beer, J., Blinov, A., Bonani, G., Hofmann, H. J. & Finkel, R. C. Use of ^{10}Be in polar ice to trace the 11-year cycle of solar activity. *Nature* **347**, 164–166 (1990).

- Lockwood, M., Stamper, R. & Wild, M. N. A doubling of the Sun's coronal magnetic field during the past 100 years. *Nature* **399**, 437–439 (1999).
- Harvey, K. L. & Zwaan, C. Properties and emergence of bipolar active regions. *Sol. Phys.* **148**, 85–118 (1993).
- Sheeley, N. R. in *The Solar Cycle* (ed. Harvey, K. L.) 1–13 (Astronomical Society of the Pacific, ASP Conf. Series Vol. 27, San Francisco, 1992).
- Wang, Y. M. & Sheeley, N. R. The rotation of photospheric magnetic fields: A random walk transport model. *Astrophys. J.* **430**, 399–412 (1994).
- Wang, Y.-M., Lean, J. & Sheeley, N. R. The long-term variation of the Sun's open magnetic flux. *Geophys. Res. Lett.* **27**, 505–508 (2000).
- Wang, Y.-M., Sheeley, N. R. & Lean, J. Understanding the evolution of the Sun's open magnetic flux. *Geophys. Res. Lett.* **27**, 621–624 (2000).
- Schrijver, C. J. *et al.* Large-scale coronal heating by the small-scale magnetic field of the Sun. *Nature* **394**, 152–154 (1998).
- Howard, R. & Labonte, B. J. Surface magnetic fields during the solar activity cycle. *Sol. Phys.* **74**, 131–145 (1981).
- Harvey, K. L. in *Solar Surface Magnetism* (eds Rutten, R. J. & Schrijver, C. J.) 347–363 (Kluwer, Dordrecht, 1994).
- Livingston, W. C., Harvey, J., Slaughter, C. & Trumbo, D. Solar magnetograph employing integrated diode arrays. *Appl. Opt.* **15**, 40–52 (1976).
- Schrijver, C. J. & Harvey, K. L. The photospheric magnetic flux budget. *Sol. Phys.* **150**, 1–18 (1994).
- Chapman, G. A., Cookson, A. M. & Dobias, J. J. Solar variability and the relation of facular to sunspot areas during solar cycle 22. *Astrophys. J.* **482**, 541–545 (1997).
- Fligge, M., Solanki, S. K., Unruh, Y. C., Fröhlich, C. & Wehrli, C. A model of solar total and spectral irradiance variations. *Astron. Astrophys.* **355**, 709–718 (1998).
- Dicke, R. H. Solar luminosity and the sunspot cycle. *Nature* **280**, 24–27 (1979).
- Hoyng, P. Is the solar cycle timed by a clock? *Sol. Phys.* **169**, 253–264 (1996).
- Wang, Y. M. & Sheeley, N. R. Solar implications of ULYSSES interplanetary field measurements. *Astrophys. J.* **447**, L143–L146 (1995).
- Wilson, P. R., Altrick, R. C., Harvey, K. L., Martin, S. F. & Snodgrass, H. B. The extended solar activity cycle. *Nature* **333**, 748–750 (1988).
- Harvey, K. L. in *The Solar Cycle* (ed. Harvey, K. L.) 335–367 (Astronomical Society of the Pacific, ASP Conf. Series Vol. 27, San Francisco, 1992).
- Friis-Christensen, E. & Lassen, K. Length of the solar cycle: An indicator of solar activity closely associated with climate. *Science* **254**, 698–700 (1991).
- Svensmark, H. & Friis-Christensen, E. Variation of cosmic ray flux and global cloud coverage—a missing link in solar-climate relationships. *J. Atmos. Terr. Phys.* **59**, 1225–1232 (1997).
- Svensmark, H. Influence of cosmic rays on Earth's climate. *Phys. Rev. Lett.* **81**, 5027–5030 (1998).
- Scherrer, P. H. *et al.* The Solar Oscillations Investigation—Michelson Doppler Imager. *Sol. Phys.* **162**, 129–188 (1995).

Acknowledgements

J. Beer and M. Lockwood provided the ^{10}Be record and the record of the reconstructed interplanetary magnetic field, respectively. We are grateful to K. Schrijver for comments on this paper.

Correspondence and requests for materials should be addressed to S.K.S. (e-mail: solanki@lindpi.mpg.de).

Coexistence of ferromagnetism and metallic conductivity in a molecule-based layered compound

Eugenio Coronado*, José R. Galán-Mascarós*†, Carlos J. Gómez-García* & Vladimir Laukhin*†

* Instituto de Ciencia Molecular, Universidad de Valencia, Dr. Moliner 50, 46100 Burjassot, Spain

Crystal engineering—the planning and construction of crystalline supramolecular architectures from modular building blocks—permits the rational design of functional molecular materials that exhibit technologically useful behaviour such as conductivity and superconductivity¹, ferromagnetism² and nonlinear optical properties³. Because the presence of two cooperative properties in the same crystal lattice might result in new physical

phenomena and novel applications, a particularly attractive goal is the design of molecular materials with two properties that are difficult or impossible to combine in a conventional inorganic solid with a continuous lattice. A promising strategy for creating this type of ‘bi-functionality’ targets hybrid organic/inorganic crystals comprising two functional sub-lattices exhibiting distinct properties. In this way, the organic π -electron donor bis(ethylenedithio)tetrathiafulvalene (BEDT-TTF) and its derivatives, which form the basis of most known molecular conductors and superconductors¹, have been combined with molecular magnetic anions, yielding predominantly materials with conventional semiconducting or conducting properties^{4,5}, but also systems that are both superconducting and paramagnetic^{6,7}. But interesting bulk magnetic properties fail to develop, owing to the discrete nature of the inorganic anions. Another strategy for achieving cooperative magnetism involves insertion of functional bulky cations into a polymeric magnetic anion, such as the bimetallic oxalato complex $[\text{M}^{\text{II}}\text{Cr}^{\text{III}}(\text{C}_2\text{O}_4)_3]^-$, but only insoluble powders have been obtained in most cases^{8–12}. Here we report the synthesis of single crystals formed by infinite sheets of this magnetic coordination polymer interleaved with layers of conducting BEDT-TTF cations, and show that this molecule-based compound displays ferromagnetism and metallic conductivity.

The bimetallic complexes $[\text{M}^{\text{II}}\text{Cr}^{\text{III}}(\text{C}_2\text{O}_4)_3]^-$ (where M^{II} is Mn, Fe, Ni, Co, Cu; M^{III} is Cr, Fe, Ru; and $(\text{C}_2\text{O}_4)^{2-}$ is the oxalate dianion) form infinite layers of oxalate-bridged hexagonal networks in the presence of bulky organic cations of the type $[\text{XR}_4]^+$ (where X is N, P; and R is Ph, *n*-Pr, *n*-Bu), which act as templates for the formation of the net (Fig. 1a)¹⁰. The magnetic properties of these coordination polymers are of interest, as they behave as ferro-, ferri- or canted antiferromagnets with critical temperatures ranging from 5 to 44 K (refs 11, 12). Moreover, it is possible to replace the electronically ‘inactive’ organic cations by paramagnetic organometallic cations⁸, metal complexes exhibiting spin crossover, or organic dyes that give rise to nonlinear optical properties⁹.

We prepared crystals of composition $[\text{BEDT-TTF}]_3[\text{MnCr}(\text{C}_2\text{O}_4)_3]$ by electrocrystallization of a methanol/benzonitrile/dichloromethane solution containing the tris-oxalate Cr(III) complex (0.01 M) and the Mn(II) ion in a ratio 2:3 and a suspension of BEDT-TTF, which was slowly oxidized at a constant current of

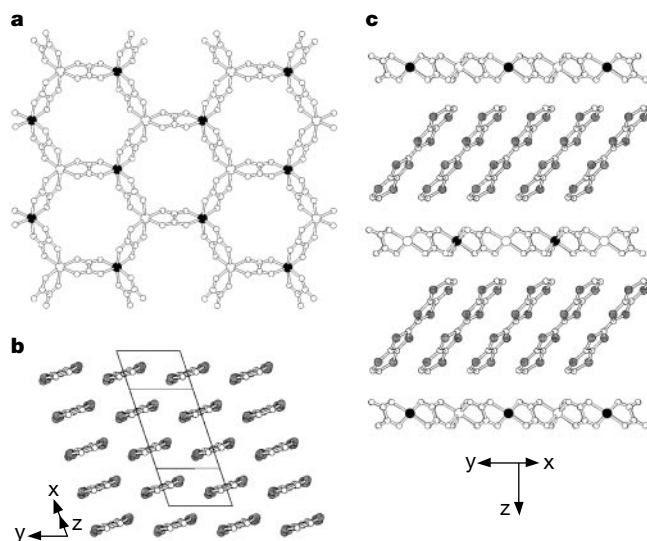


Figure 1 Structures of the hybrid material and the two sublattices. **a**, View of the $[\text{M}^{\text{II}}\text{M}^{\text{III}}(\text{C}_2\text{O}_4)_3]^-$ bimetallic layers. Filled and open circles in the vertices of the hexagons represent the two types of metals. **b**, Structure of the organic layer, showing the β packing of the BEDT-TTF molecules. **c**, Representation of the hybrid structure along the *c* axis, showing the alternating organic/inorganic layers.

† Present addresses: Department of Chemistry, Texas A&M University, College Station, Texas, USA (J.R.G.-M.); ICMB-CSIC, Campus de la UAB, 08193 Bellaterra, Spain (V.L.).

0.1 μA . After a week, thin, shiny brown plates of this salt were collected from the platinum electrode. The single-crystal X-ray data obtained at 120 K indicate a strong triclinic sublattice, P-1, with $a = 5.17(2)$ Å, $b = 6.38(2)$ Å, $c = 16.61(6)$ Å, $\alpha = 87.15(5)^\circ$, $\beta = 85.12(5)^\circ$ and $\gamma = 66.60(5)^\circ$; $V = 501(3)$ Å³. The final refinement with 1,317 reflections / 151 parameters ($I > 4\sigma$) gave $R_1 = 7.5\%$ ($R_w = 18.77$). The structure consists of organic layers of BEDT-TTF cations alternating with honeycomb layers of the bimetallic oxalato complex $[\text{MnCr}(\text{C}_2\text{O}_4)_3]^-$ (Fig. 1). The BEDT-TTF cations are tilted with respect to the inorganic layer by an angle of 45° . The layers of BEDT-TTF are constructed by only one crystallographically independent BEDT-TTF molecule located with the central carbon-carbon double bond on an inversion centre. In this model all donor molecules are related by translations, and adopt pure two-dimensional pseudohexagonal β -type packing. Thus, a precise structure of the organic sublattice can be obtained. From the lengths of the C = C and C-S bonds, we estimate a positive charge on the BEDT-TTF molecule of +0.34, in agreement with the chemical formula. In this structural model, the inorganic oxalate-based bimetallic layer is crystallographically disordered. This type of disorder is frequently encountered in layered systems. In the present case it may arise from the presence of stacking faults similar to those found in other charge-transfer salts containing polymeric anions¹³, and in related $\text{A}[\text{M}^{\text{II}}\text{M}^{\text{III}}(\text{C}_2\text{O}_4)_3]$ materials¹⁴. In fact, the positions assigned to the metal ions, which define the central plane of the inorganic layer, appear as linear arrays in the structure; this indicates that the stacking faults are due to translations of the honeycomb layers along the vector defined by one of the diagonals of the hexagon.

The magnetic properties of this compound show that it is a magnet below a critical temperature, T_c , of 5.5 K. In particular, a.c. magnetic measurements exhibit a sharp peak in the in-phase signal, accompanied by an out-of-phase signal below T_c (Fig. 2a). The

ferromagnetic nature of the transition is confirmed by the field dependence of the isothermal magnetization performed at 2 K, which shows a rapid saturation of the magnetization, M_s , with a value of $7.1 \mu_B$ (Fig. 2b). This value is slightly smaller than that expected ($\sim 8 \mu_B$) and, as already noted¹¹, arises from the presence of a spin canted structure in the ferromagnetic state. The hysteretic behaviour shows quite small coercive fields ($H_c \approx 5\text{--}10$ Oe) at low temperatures, so this compound is a soft ferromagnet. All the above magnetic features are identical to those already reported for the other $\text{Mn}(\text{II})\text{Cr}(\text{III})$ -oxalato layered magnets ($[\text{NBu}_4]^+$ (ref. 11) and $[\text{CoCp}^*_2]^+$ (ref. 8) salts, where Cp^* is pentamethylcyclopentadienyl). Hence, the magnetic behaviour constitutes an unambiguous confirmation of the presence of the honeycomb bimetallic network in

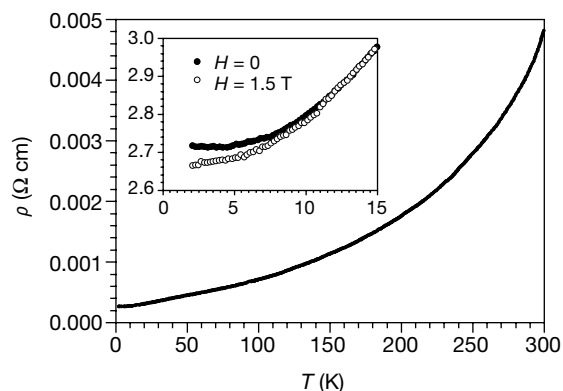


Figure 3 Thermal variation of the electrical resistivity in the hybrid material. Inset, the low-temperature magnetoresistance with the magnetic field applied perpendicular to the layers (open circles); the ρ values are multiplied by a factor of 10^4 .

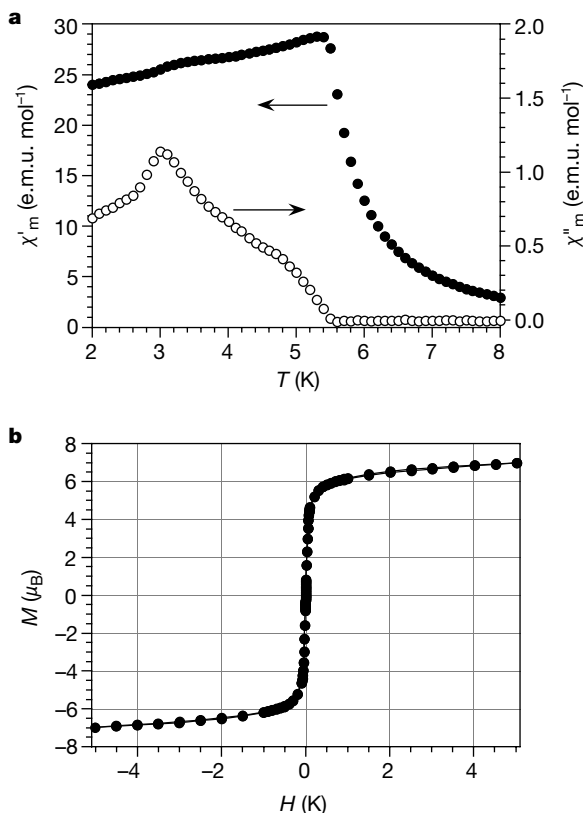


Figure 2 Magnetic characterization of the hybrid material. **a**, In-phase (filled circles) and out-of-phase (open circles) components of the a.c. susceptibilities at 110 Hz. **b**, Plot of magnetization versus field at 2 K.

the BEDT-TTF derivative. Furthermore, the properties indicate that the larger interlayer separation imposed by the insertion of BEDT-TTF into the bimetallic oxalato layers (16.61 Å compared to 9–10 Å in the other layered magnets) has no effect on the magnetic properties. This establishes that the magnetic ordering in these two-dimensional phases occurs within the bimetallic layers, and is triggered by the small magnetic anisotropy of the metallic ions¹⁵.

The transport properties of the material are depicted in Fig. 3. We performed the resistance measurements on the plate-like single crystals using the standard four-contact d.c. method. The current was flowing in the plane of the layers. The conductivity at room temperature reaches a value as high as $\sim 250 \text{ S cm}^{-1}$. The temperature dependence of the resistance exhibits a metallic behaviour over the entire temperature range that we investigated (down to 2 K). Depending on the sample quality, the resistivity decreases by factors of about 15–25 as the temperature is decreased from 300 to 2 K. Application of a magnetic field (up to 5 T) perpendicular to the layers leads to the appearance of a negative magnetoresistance at temperatures below about 10 K (Fig. 3 inset), whereas the resistance is practically field-independent when the field is applied parallel to the layers. This feature may be caused by the presence of the internal fields generated at low temperatures by the ferromagnetic inorganic layers. We note that this is the only evident interplay between the two kinds of sublattices (ferromagnetic and conducting). No influence of the conducting sublattice on the magnetic properties has been encountered. Thus, we can say that from the electronic point of view the two sublattices are quasi-independent. This feature, characteristic of this class of molecular materials, is in sharp contrast with what is found in the classical ferromagnetic metals like iron or gadolinium, where the two electronic sublattices are strongly interacting and cannot be decoupled.

Further characterization of the material is in progress, involving measurements of resistivity and magnetoresistance at lower temperatures and in high fields, and under pressure. Moreover, to establish whether the lack of superconductivity above 2 K is intrinsic to the organic sublattice or due to the presence of the ferromagnetic oxalato layer, we are exploring the preparation of isostructural salts with a paramagnetic or even diamagnetic bimetallic oxalato layer, facilitated by the versatility of this crystal lattice. Another avenue of investigation could be the possible insertion of other organic donors that are similar to BEDT-TTF (for example, the selenium and oxygen derivatives). □

Received 4 May; accepted 5 October 2000.

- Williams, J. M. *et al.* Organic superconductors—new benchmarks. *Science* **252**, 1501–1508 (1991).
- Miller, J. S. & Epstein, A. J. Organic and organometallic magnetic materials—designer magnets. *Angew. Chem. Int. Edn Engl.* **33**, 385–415 (1994).
- Zyss, J. (ed.) *Molecular Nonlinear Optics* (Academic, New York, 1994).
- Day, P. *et al.* Structure and properties of tris[bis(ethylenedithio)-tetrathiafulvalenium]tetrachlorocopper(II) hydrate, (BEDT-TTF)₃CuCl₄·H₂O: first evidence for coexistence of localized and conduction electrons in a metallic charge-transfer salt. *J. Am. Chem. Soc.* **114**, 10722–10729 (1992).
- Coronado, E. *et al.* Magnetic molecular metals based on the organic donor BET-TTF (BET-TTF bis(ethylenedithio)tetrathiafulvalene). *Adv. Mater.* **9**, 984–987 (1997).
- Kurmoo, M. *et al.* Superconducting and semiconducting magnetic charge transfer salts: (BEDT-TTF)₄Fe(C₂O₄)₃C₆H₅CN (A = H₂O, K, NH₄). *J. Am. Chem. Soc.* **117**, 12209–12217 (1995).
- Kobayashi, H. *et al.* New BETS conductors with magnetic anions (BETS = bis(ethylenedithio)tetrathiafulvalene). *J. Am. Chem. Soc.* **118**, 368–377 (1996).
- Coronado, E. *et al.* Hybrid molecular magnets obtained by insertion of decamethylmetalocenium in layered bimetallic oxalate complexes. Syntheses, structure and magnetic properties of the series [Z^{III}Cp⁺][M^{II}M^{III}(ox)₃] (Z^{III} = Co, Fe; M^{III} = Cr, Fe; M^{II} = Mn, Fe, Co, Cu, Zn; Cp⁺ = pentamethylcyclopentadienyl). *Chem. Eur. J.* **6**, 552–563 (2000).
- Coronado, E. *et al.* Hybrid molecular materials formed by two molecular networks. Towards multiproperty materials. *Mol. Cryst. Liq. Cryst.* **334**, 679–691 (1999).
- Decurtins, S. *et al.* in *Magnetism: A Supramolecular Function* (ed. Kahn, O.) 487–508 (NATO ASI Ser. C484, Kluwer Academic, 1996).
- Tamaki, H. *et al.* Design of metal-complex magnets. Syntheses and magnetic properties of mixed-metal assemblies {[NBu₄][M(Cr(ox)₃)]} (NBu₄⁺ = tetra(n-butyl)ammonium ion; ox²⁻ = oxalate ion; M = Mn²⁺, Fe²⁺, Co²⁺, Ni²⁺, Cu²⁺, Zn²⁺). *J. Am. Chem. Soc.* **114**, 6974–6979 (1992).
- Mathonière, C. *et al.* Ferrimagnetic mixed-valency and mixed-metal tris(oxalato) iron (III) compounds: synthesis, structure and magnetism. *Inorg. Chem.* **35**, 1201–1206 (1996).
- Beno, M. A. *et al.* The first ambient pressure organic superconductor containing oxygen in the donor molecule, β_m-(BEDO-TTF)₃Cu₂(NCS)₃, T_c = 1.06 K. *Inorg. Chem.* **29**, 1599–1601 (1990).

- Nuttall, C. J. & Day, P. Modeling stacking faults in the layered molecular-based magnets AM^{II}Fe(C₂O₄)₃ M^{II} = Mn, Fe; A = organic cation. *J. Solid State Chem.* **147**, 3–10 (1999).
- Antorrena, G. *et al.* Thermal and magnetic study of mixed-metal oxalate-bridged 2d magnets. *J. Magn. Mater.* **196–197**, 581–583 (1999).

Acknowledgements

V.L. is on leave from the Institute of Problems of Chemical Physics, 142432 Chernogolovka, Russia. We thank K. R. Dunbar for access to X-ray equipment at Texas A&M University and for discussions. This work was supported by the Spanish Ministerio de Ciencia y Tecnología and by the European Union (Network on Molecular Magnetism: From Materials to Devices).

Correspondence and requests for materials should be addressed to E.C. (e-mail: eugenio.coronado@uv.es).

Direct imaging of the pores and cages of three-dimensional mesoporous materials

Yasuhiro Sakamoto*, Mizue Kaneda*, Osamu Terasaki*†‡, Dong Yuan Zhao‡, Ji Man Kim§, Galen Stucky||, Hyun June Shin¶ & Ryong Ryoo¶#

* Department of Physics and CIR, Tohoku University, Sendai 980-8578, Japan

† CREST, Japan Science and Technology Corporation, Tohoku University, Sendai 980-8578, Japan

‡ Department of Chemistry, Fudan University, Shanghai 200433, China

§ Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, USA, and Catalysis Center for Molecular Engineering, Korea Research Institute of Chemical Technology, Yusong, Taejeon, 305-600, Korea

|| Department of Chemistry and Biochemistry and Materials Department, University of California, Santa Barbara, California 93106, USA

¶ Materials Chemistry Laboratory, Department of Chemistry (School of Molecular Science-BK21), Korea Advanced Institute of Science and Technology, Taejeon 305-701, Korea

These authors contributed equally to this work

Mesostuctured composite materials, with features ranging from 20 to 500 Å in size, are obtained by the kinetically controlled competitive assembly of organic and inorganic species into nanostructured domains. Short-range order is limited, and long-range order is determined by weak forces such as van der Waals or hydrogen-bonding. Three-dimensional mesoporous materials obtained by removing the organic phase are of particular interest for applications such as catalysis and chemical sensing or separation, for which structural features such as cavity shape, connectivity and ordered bimodal porosity are critical. But atomic-scale structural characterization by the usual diffraction techniques is challenging for these partially ordered materials because of the difficulty in obtaining large (> 10 μm) single crystals, and because large repeat spacings cause diffraction intensities to fall off rapidly with scattering angle so that only limited small-angle data are available. Here we present a general approach for the direct determination of three-dimensional mesoporous structures by electron microscopy. The structure solutions are obtained uniquely without pre-assumed models or parametrization. We report high-resolution details of cage and pore structures of periodically ordered mesoporous materials^{1,2}, which reveal a highly ordered dual micro- and mesoscale pore structure.

It is not straightforward to obtain structural information from high-resolution electron microscopy (HREM) images for three-dimensional (3D) mesoporous materials, as the images are pro-

jected structural information of the materials along the direction of incidence. To date, the structures of the pores and cages defined by the inorganic domain structure have been inferred indirectly from absorption isotherm measurements or by using assumed structural models to generate diffraction patterns or images to compare with

experimental data^{3,4}. The imprecise results of this approach, particularly for X-ray or neutron scattering⁵, along with the kinetic nature of the synthesis, have made it difficult to define these materials, as wide variations in structure and order of materials from apparently the same synthesis procedure may not be resolved.

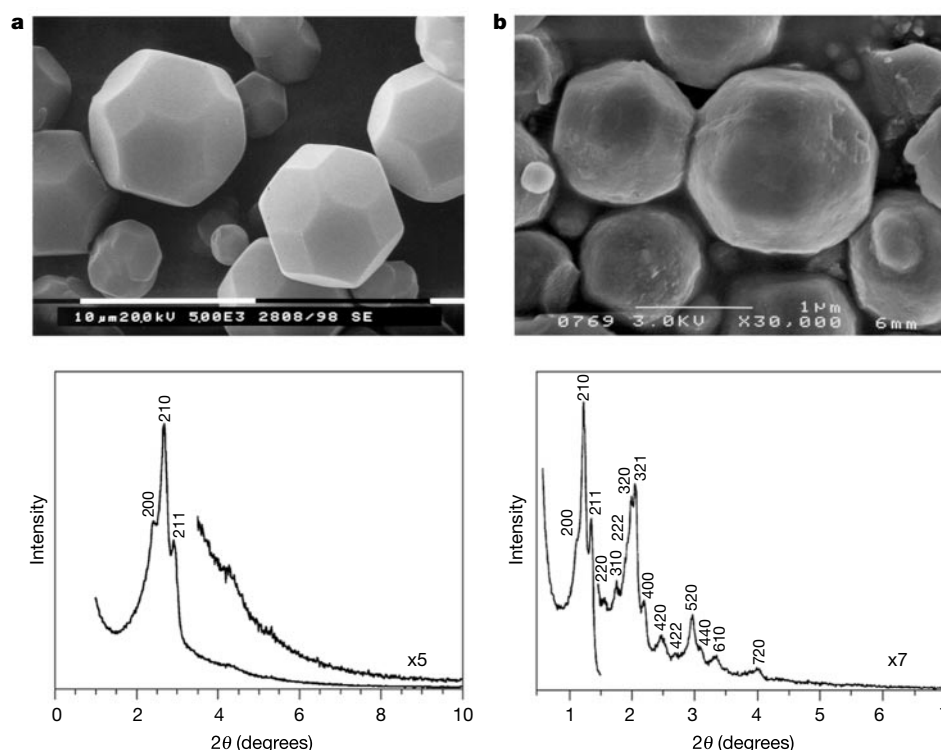


Figure 1 SEM images and X-ray powder diffraction patterns of as-synthesized SBA-1 (**a**) and SBA-6 (**b**). The crystal morphologies are consistent with the cubic point group symmetry of $m\bar{3}m$. The X-ray diffraction patterns for these samples are indexed as $Pm\bar{3}n$

with $a = 86 \text{ Å}$ (as-synthesized), 73 Å (calcined) for SBA-1, and $a = 160 \text{ Å}$ (as-synthesized), 146 Å (calcined) for SBA-6. Unit cell constants for SBA-1 and SBA-6 decrease 15% and 9%, respectively, after calcination.

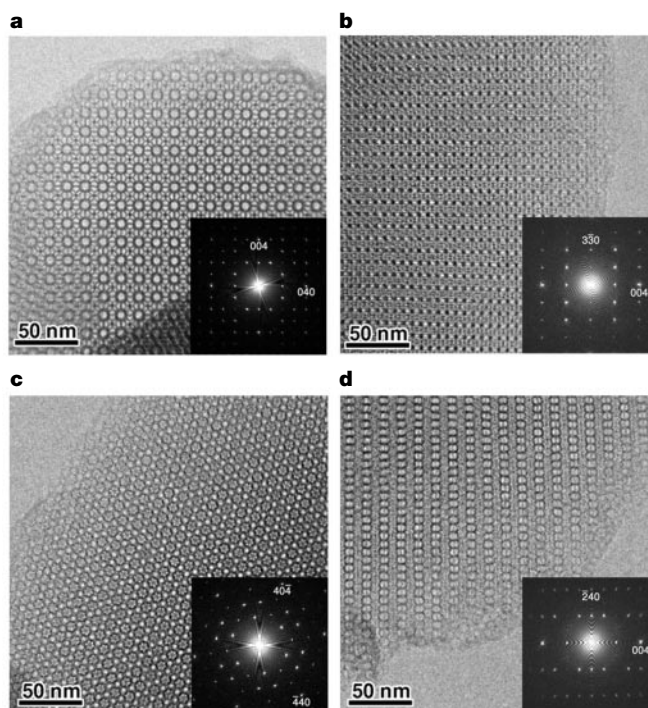


Figure 2 HREM images of SBA-6 (calcined) and their Fourier diffractograms. **a**, [100]; **b**, [110]; **c**, [111]; and **d**, [210]. The images were taken by a JEM-3010 instrument (JEOL)

(spherical aberration constant $C_s = 0.6 \text{ mm}$, chromatic aberration constant $C_c = 1.3 \text{ mm}$, operated at 300 kV) equipped with slow-scan CCD camera.

Among the 3D mesoporous materials, SBA-1^{6,7} and a mesocomposite phase that we report here (SBA-6), synthesized in acidic and basic conditions respectively, give X-ray powder diffraction patterns that can be indexed in the cubic system and which are similar to the diffraction patterns observed for cubic $Pm\bar{3}n$ globular arrays obtained in amphiphilic surfactant solution^{8,9}. Another system, SBA-16 synthesized in acidic conditions with the F127 triblock copolymer, has been reported to be $Im\bar{3}m$ on the basis of X-ray and some preliminary electron-microscopy observations^{1,2}. Confidence in the assignment of space groups to SBA-1 and SBA-16 was limited by the available data. Furthermore, space group symmetry is not structure and the resolved structures of 3D mesoporous cage materials have not been determined.

Mesoporous silica materials have two structural characteristics: disorder on the atomic scale (short-range) but distinct order on the mesoscopic scale (long-range). The self-organization of surfactants is an important mechanism for producing the periodic structure on the mesoscale, and this organization is highly sensitive to changes in both local and average structures on the atomic scale—it depends on synthesis conditions such as composition, temperature and pH. From a very small crystal, we can, by electron microscopy, obtain single-crystal structural information, free from the local variations that typically contribute to both a small number of reflections and large peak widths in X-ray diffraction experiments. We have developed an approach that uses electron crystallography to solve 3D structures of mesoporous materials with disordered wall structures. The 3D structural solution makes clear, at the nanoscale level, the sizes and shapes of the pores and cages, their arrangements and their connectivity, including sizes of openings.

Our method makes use of high-resolution electron microscopy (HREM), and was developed especially for periodic structural arrangements with mesoscale ordering but not necessarily short-range ordering. As diffraction intensities can give only moduli of the structure factors, while HREM images give both the amplitudes and phases of the necessary Fourier coefficients, the data collection and analysis begins with obtaining two-dimensional (2D) projections of the image of the 3D structure of the sample. Fourier-transform analysis of these 2D images collectively gives a 3D representation of the structure in reciprocal space. Once this set of structure factors with known amplitudes and phases is obtained in 3D reciprocal space, the 3D real space structure can be obtained in a straightforward way through the inverse Fourier transformation. The methodology has been successfully applied to a variety of mesoporous silica SBA structures^{1,2}. Details are given here for three examples; the above-mentioned SBA-1, SBA-6 and SBA-16. Structural details for SBA-2, SBA-8, SBA-11 and several previously unreported phases will be described elsewhere.

SBA-1 and SBA-16 were synthesized following procedures based on those described previously, under acidic conditions, with a pH below the aqueous isoelectric point of silica^{1,2}. An improved synthesis procedure was used for SBA-1, which differs from the one originally described^{6,7} in that the synthesis was carried out at 0 °C over a period of one week¹⁰. The synthesis of SBA-6 has not been previously reported, and this is (to our knowledge) the first example of a cubic cage structure synthesized under basic synthesis conditions (see Methods).

The as-made (acid-synthesized) SBA-1 contains one anion for every charged surfactant with a formally neutral framework, whereas in (base-synthesized) SBA-6 the walls of the silica framework are charged to neutralize the amphiphilic gemini surfactant charge; thus the mesostructure reaction product compositions for SBA-1 and SBA-6 are distinctly different. The silica organization in the walls of the two structures differ for acid-catalysed (SBA-1) and base-catalysed (SBA-6) silica hydrolysis and polymerization¹¹, and the acid-polymerized silica phase has walls that are 2 to 4 times thicker because of the reduced framework charge during silica condensation.

Figure 1a and b show scanning electron microscope (SEM) images that are typical for SBA-1 and SBA-6. Both materials are crystalline at the mesoscale level, but have glass-like disorder in the walls. The HREM images are shown for SBA-6 with [100], [110], [111] and [210] incidences in Fig. 2. In order to enhance the contribution to image formation from the reflections with low-angle scattering vectors, the HREM images shown here were taken not at Scherzer focus but at underfocus ($\sim 2,000$ Å). Corresponding Fourier diffractograms (moduli of Fourier transforms) are shown as insets, and they clearly show extinction conditions for the reflections. Fourier diffractograms from thin regions close to the edge of the crystal avoid the usual difficulties associated with multiple scattering contributions in electron diffraction patterns. Because of the high degree of order on the mesoscopic scale, we have confirmed independently that if the specimen is thicker than 300 Å (at operating conditions of 300 kV), multiple scattering effects become discernible even in these atomistically disordered materials, so that it is difficult to obtain the conditions for reflection solely from the electron diffraction patterns. These data for SBA-1 and SBA-6 clearly show that either $Pm\bar{3}n$ (no. 223) or $P43n$ (no. 218) is an allowed space group. However we conclude that the space group is $Pm\bar{3}n$, because the crystal morphologies (Fig. 1a, b) show

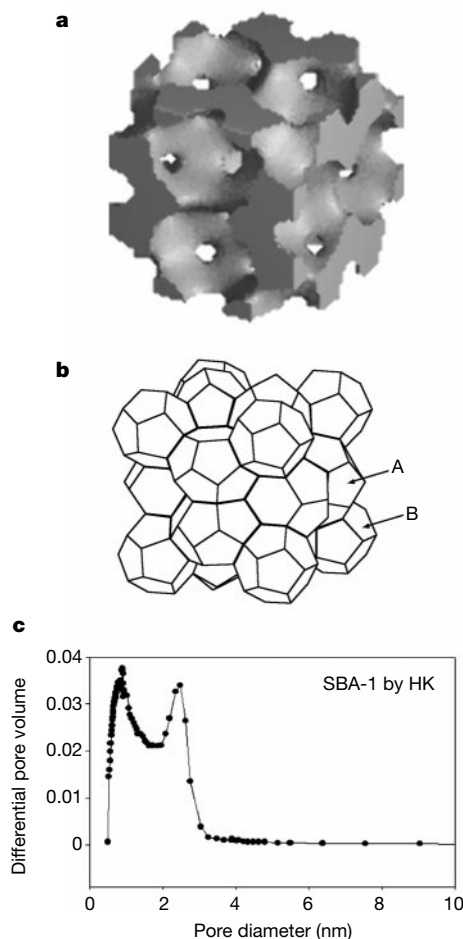


Figure 3 Direct image of 3D cages and bimodal pore structure in SBA-6 and SBA-1. **a**, 3D structure of SBA-6 obtained from the electrostatic potential map, showing large A-cages and small B-cages with associated large and small pores. **b**, Schematic drawing of the arrangement of A- and B-cages in A_3B type structures, where the B-cage is the smaller, with a diameter of 73 Å (SBA-6) at (0,0,0) and (1/2, 1/2, 1/2), and the A-cage is the larger, with a diameter of ~ 85 Å (SBA-6) at (1/2, 0, 1/4), (1/2, 0, 3/4), (0, 1/4, 1/2), (0, 3/4, 1/2), (1/4, 1/2, 0) and (3/4, 1/2, 0). For SBA-1, the diameters of the A and B cages are 40 Å and 33 Å, respectively. **c**, Pore sizes for the calcined SBA-1 framework determined by Horvath-Kawazoe analysis of the argon adsorption isotherm branch.

point group symmetries of $m\bar{3}m$, whereas $P43n$ should have a point symmetry of $\bar{4}3m$.

The crystal structure factors are obtained from a set of Fourier transforms of the HREM images by correcting for the effect of an objective lens contrast transfer function (CTF). Taking the origin at the inversion centre ($m\bar{3}$), all crystal structure factors become real, that is, the phases of the structure factors are 0 or π (see Supplementary Information). At the underfocus condition used, CTF is almost linear with respect to the magnitude of the scattering vector, q , for reflections important for the Fourier sum. Therefore, for the spatial resolution that is accessible from the available diffraction data, the result is almost independent of the focus conditions. This gives 42 independent reflections out of 44 reflections with larger spacing than 20 Å. The intensities of the remaining two reflections are weak in the electron diffraction experiment.

The 3D electrostatic potential distribution for SBA-6 is obtained from the Fourier summation of the crystal structure factors (these structure factors are available as Supplementary Information). Using the pore volume of $V_p = 0.86 \text{ cm}^3 \text{ g}^{-1}$ obtained from N_2 gas adsorption data and the wall density of $D_w = 2.2 \text{ g cm}^{-3}$ determined by helium pycnometry, a threshold in the potential density is determined for differentiating between the amorphous wall and enclosed cavities, giving the 3D structure shown in Fig. 3. A-cages and B-cages, differing in size, are arranged as in an MEP clathrate type (A_3B) structure (Fig. 3b)¹².

The HREM analysis used here reveals an unusual aspect of the structures of SBA-1 and SBA-6, namely bimodal (micro-meso) porosity. In the structure of SBA-6, a B-cage is surrounded by 12 A-cages that are connected through mesoporous openings of 20 Å, while the openings between A-cages are about 33×41 Å. By the

same procedure, the 3D electrostatic potential distribution and structure of SBA-1 have been determined. In this case, $V_p = 0.60 \text{ cm}^3 \text{ g}^{-1}$, $D_w = 2.0 \text{ g cm}^{-3}$, with a greater average wall thickness based on sorption data. The mesopore size for SBA-1 determined by electron microscopy is 15×22 Å while the micropore size is 2 Å. Horvath-Kawazoe analysis of the adsorption-desorption isotherm (BET) data for SBA-1 (Fig. 3c) gives pore sizes of 24 Å and 8 Å (see Methods). There is no doubt about the bimodal porosities of these materials, although further work is needed to resolve the differences in pore diameters as determined by the BET and the HREM techniques. 3D HREM analysis gives details on asymmetric pore sizes and pore configuration that cannot be obtained by other techniques. We note that the pores of SBA-1 and SBA-6 can be made different sizes by using the appropriate surfactants.

We now apply our technique to SBA-16. Figure 4 shows HREM images for [100], [110] and [111] incidences and the corresponding Fourier diffractograms of the images from thin areas. From the observed extinction ($h + k + l = 2n + 1$), four space groups, $Im\bar{3}$, $I432$, $I43m$, and $Im\bar{3}m$, are possible. The morphology of the observed crystals of this material is simple cubic, indicating a point group of $m\bar{3}m$ and confirming that $Im\bar{3}m$ ($a = 133$ Å) is a suitable space group. Following the procedure described above, the 3D structure was determined and is shown in Fig. 5. Spherical cavities of 95 Å diameter are arranged in a body-centred-cubic array, and the cavities are connected through mesoporous openings of 23 Å along the [111] directions. We note that the surface with $(\cos\pi x \times \cos\pi y) + (\cos\pi x \times \cos\pi z) + (\cos\pi y \times \cos\pi z) = 0$ (Fig. 5c), which is related to the minimal surface of I-WP (body centred, wrapped package), describes well the configuration of SBA-16.

The methodology described here opens a way to solve structures

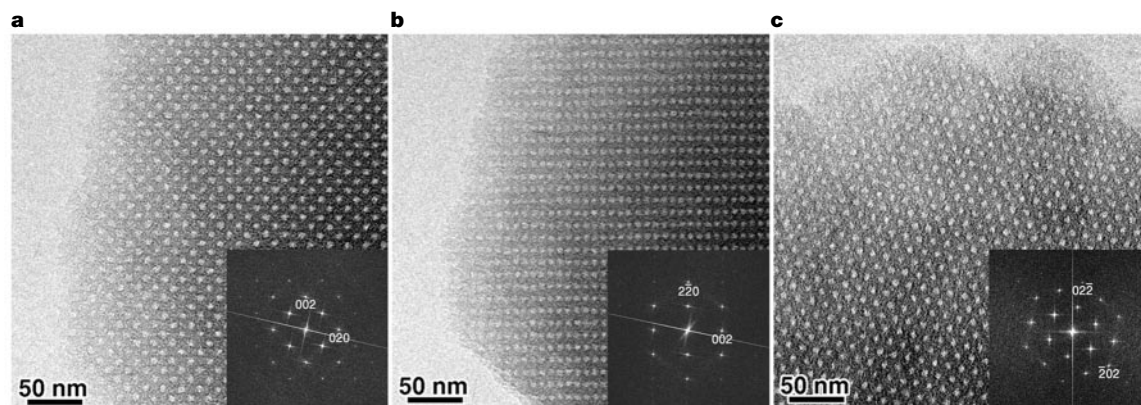


Figure 4 HREM images of SBA-16 (calcined) and their Fourier diffractograms. **a**, [100]; **b**, [110]; **c**, [111]. The images were taken by a JEM-1250 instrument (JEOL) ($C_s = 1.7$ mm, $C_c = 4.0$ mm, operated at 1,250 kV) with imaging plates.

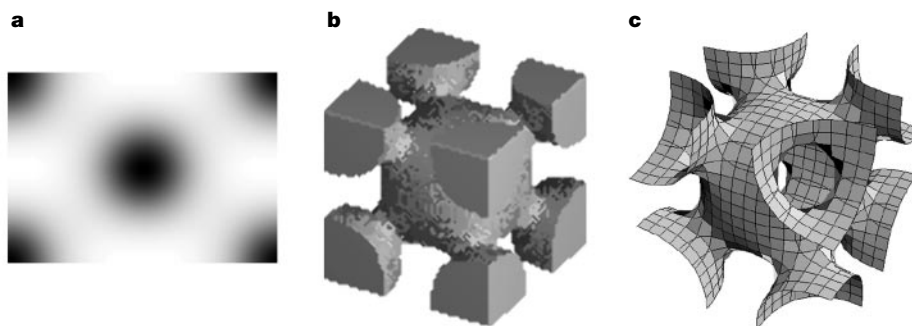


Figure 5 Direct image of 3D pore structure of SBA-16. **a**, Electrostatic potential map of SBA-16 parallel to (110) through the centre of the cell. **b**, 3D arrangement of a cavity and its interconnection. Black corresponds to the cavity. **c**, Mean cavity surface for SBA-16.

of 3D periodic mesostructured materials without assuming any structural models. The resolution for the structure is primarily limited by the quality of the HREM images, which depends on the long-range mesoscale ordering. Therefore, although further progress may give better resolution, we expect no future change to the present conclusions about the structures of SBA-1, SBA-6 and SBA-16, because the validity of the solutions does not depend on the resolution. This is a characteristic of our method that makes it different from other approaches. We also suggest that the results presented here provide a quantitative topological description of ordered mesostructured composites, and that such descriptions are essential in understanding the properties and possible applications of the composites. The resolution of periodically ordered, 3D arrangements of bimodal (meso-micro) pores in SBA-1 and SBA-6 makes it possible to consider the detailed characterization of the range of complicated porous phases that are now synthetically achievable. □

Methods

Synthesis of SBA-6

3.75 g of tetraethoxysilane (TEOS) was added with magnetic stirring to a clear solution containing 0.5 g of the gemini surfactant 18B₄₋₃₋₁ (N,N,N',N'-pentamethyl-N'-[4-(4-octadecyloxyphenoxy)-butyl]-propane-1,3-diammonium dibromide, C₁₈H₃₇OC₆H₄OC₆H₄N(CH₃)₂C₃H₆N(CH₃)₃Br₂), 45.4 g of doubly distilled water, and 3.69 g of benzyltrimethylammonium hydroxide at room temperature. Stirring was continued for 20 h after the addition of TEOS at room temperature. The reaction gel mixture was heated for 2 d at 80 °C without stirring. The precipitate was filtered and dried in air at room temperature.

Determination of properties

Ar adsorption and desorption isotherms were measured at 87 K. Pore volumes (cm³ g⁻¹) for SBA-1, SBA-6 and SBA-16 are 0.6, 0.86 and 0.45, respectively, and the ratios of the pore volume to unit cell are respectively 0.57, 0.65 and 0.47. The surface-area/pore-volume ratio (2.26 × 10⁹ m⁻¹) for SBA-1 is nearly three times that of SBA-6 (7.93 × 10⁸ m⁻¹). The silica wall densities determined with an AccPyc 1300 helium pycnometer are also substantially different for SBA-1 (2.00 g cm⁻³) and SBA-6 (2.20 g cm⁻³).

Received 23 May; accepted 6 October 2000.

1. Zhao, D. *et al.* Triblock copolymer syntheses of mesoporous silica with periodic 50 to 300 Ångström pores. *Science* **279**, 548–552 (1998).
2. Zhao, D., Huo, Q., Feng, J., Chmelka, B. F. & Stucky, G. D. Nonionic triblock and star diblock copolymer and oligomeric surfactant syntheses of highly ordered, hydrothermally stable, mesoporous silica structures. *J. Am. Chem. Soc.* **120**, 6024–6036 (1998).
3. Alfredsson, V. & Anderson, M. W. Structure of MCM-48 revealed by transmission electron microscopy. *Chem. Mater.* **8**, 1141–1146 (1996).
4. Monnier, A. *et al.* Cooperative formation of inorganic-organic interfaces in the synthesis of silicate mesostructures. *Science* **261**, 1299–1303 (1993).
5. Schacht, S., Janicke, M. & Schütt, F. Modeling X-ray patterns and TEM images of MCM-41. *Microporous Mesoporous Mater.* **22**, 485–493 (1998).
6. Huo, Q. *et al.* Generalized syntheses of periodic surfactant/inorganic composite materials. *Nature* **368**, 317–321 (1994).
7. Huo, Q. *et al.* Organization of organic molecules with inorganic molecular species into nanocomposite biphasic arrays. *Chem. Mater.* **6**, 1176–1191 (1994).
8. Auvray, X. *et al.* X-ray diffraction and freeze-fracture electron microscopy study of the cubic phase in the cetylpyridinium chloride formamide and cetyltrimethylammonium chloride formamide systems. *Langmuir* **9**, 444–448 (1993).
9. Charvolin, J. & Sado, J. F. Periodic systems of frustrated fluid films and “micellar” cubic structures in liquid crystals. *J. Phys. France* **49**, 521–526 (1988).
10. Ryoo, R., Kim, J. M. & Ko, C. H. in *Studies in Surface Science and Catalysis* Vol. 117 (eds Bonnevot, L., Beland, F., Danumah, C., Giasson, S. & Kaliaguine, S.) 151–158 (Elsevier, Amsterdam, 1998).
11. Nakanishi, K. Pore structure control of silica gels based on phase separation. *J. Porous Mater.* **4**, 67–112 (1997).
12. Geis, H. Studies on clathrasils. III. Crystal structure of melanophlogite, a natural clathrate compound of silica. *Z. Kristallogr.* **164**, 247–257 (1983).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported in part by CREST, Japan Science and Technology Corporation (O.T.), by the National Research Laboratory Program of Korea (R.R.), and by the National Science Foundation (G.D.S.) and the Army Research Office (G.D.S.). O.T. thanks S. Andersson for encouragement and support. Y.S. thanks the Japan Society for the Promotion of Science.

Correspondence and requests for materials should be addressed to O.M. (e-mail: terasaki@msp.phys.tohoku.ac.jp) or R.R. (e-mail: r.ryoo@mail.kaist.ac.kr).

Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data

Alexandre Ganachaud* & Carl Wunsch

MIT 54-1517, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

Through its ability to transport large amounts of heat, fresh water and nutrients, the ocean is an essential regulator of climate^{1,2}. The pathways and mechanisms of this transport and its stability are critical issues in understanding the present state of climate and the possibilities of future changes. Recently, global high-quality hydrographic data have been gathered in the World Ocean Circulation Experiment (WOCE), to obtain an accurate picture of the present circulation. Here we combine the new data from high-resolution trans-oceanic sections and current meters with climatological wind fields, biogeochemical balances and improved a priori error estimates in an inverse model, to improve estimates of the global circulation and heat fluxes. Our solution resolves globally vertical mixing across surfaces of equal density, with coefficients in the range $(3\text{--}12) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$. Net deep-water production rates amount to $(15 \pm 12) \times 10^6 \text{ m}^3 \text{ s}^{-1}$ in the North Atlantic Ocean and $(21 \pm 6) \times 10^6 \text{ m}^3 \text{ s}^{-1}$ in the Southern Ocean. Our estimates provide a new reference state for future climate studies with rigorous estimates of the uncertainties.

Obtaining a consistent picture of the oceanic circulation requires adjusting thousands of parameters consistently with a priori error estimates. We present here our best estimate from selected hydrographic data (Fig. 1), which will improve with the appearance of new data. Mass flux is the most basic element of the circulation and Fig. 2 shows the best-estimate coast-to-coast integrated water mass transports for selected density classes. A volume of $15 \pm 2 \text{ Sv}$ ($1 \text{ Sverdrup} = 1 \times 10^6 \text{ m}^3 \text{ s}^{-1}$) of North Atlantic Deep Water (NADW) is produced in the northern North Atlantic Ocean and moves southward, entraining Antarctic Bottom Water (AABW) from below, and Antarctic Intermediate Water (AAIW) from above. As a result, the NADW is increased to $23 \pm 3 \text{ Sv}$ as it exits the South Atlantic at 30° S. In the Southern Ocean, a total of $21 \pm 6 \text{ Sv}$ of bottom water is formed from lower Circumpolar Deep Water (CDW)—which corresponds approximately to the lower NADW density range. Bottom water inflows (NADW + AABW mixture) to the Atlantic, Indian and Pacific oceans are $6 \pm 1.3 \text{ Sv}$, $11 \pm 4 \text{ Sv}$ and $7 \pm 2 \text{ Sv}$, respectively. In the Indian and Pacific oceans, most of this water returns southward at deep and intermediate levels. These net values are the sums of large, strongly spatially varying, flows of opposing sign, and thus oversimplify the actual circulation; a detailed description of the circulation within each ocean basin will be published elsewhere^{3,4}. Our standard model estimate of the inflow in the South Pacific Ocean is in the lower range of previously published values, but it depends directly upon the weight given to the “PO” phosphate–oxygen combination (see Methods^{4,5}) conservation constraints relative to mass conservation³. The deep inflow to the North Pacific Ocean is also weaker than previously found³, as a consequence of our consideration of heat and salt conservation in the northern parts of those basins.

No definition of bottom-water formation can be completely unambiguous because of the entrainment of ambient fluid during the sinking process. In our Southern Ocean definition, the bottom-

* Present address: Laboratoire de Physique des Océans, IFREMER, 29280 Plouzané, France.

water formation rate corresponds to the total amount of water crossing the neutral surface, $\gamma^n = 28.11 \text{ kg m}^{-3}$ downwards, and subsequently ventilating the deep Atlantic, Indian and Pacific basins. This water is provided primarily from the density range of lower NADW, so that the two sources of deep water are not additive, as is often assumed in qualitative calculations⁶. Consistent with conventional wisdom, no very large amount of deep water forms in the North Pacific or Indian oceans, although intermediate waters are produced there. Transient tracer measurements suggest AABW formation rates of water denser than $\gamma^n = 28.27 \text{ kg m}^{-3}$ at 5 to 8 Sv (ref. 7). Because this density class is almost immediately mixed with CDW above it when exiting the Southern Ocean, it is not separately resolved in our model, and there is no inconsistency. However, there is a contradiction with the suggestion that the current production rate of AABW recently decreased⁶, as we find no inconsistency between the deep, contemporary, geostrophic circulation and the interior tracer field which reflects the circulation integrated over hundreds of years. We believe that the origin of the disagreement is in those authors' *ad hoc* assumptions⁶ about the originating water masses.

The integrated circulation in the upper layers is more uncertain (Fig. 2, red arrows) owing to the enhanced temporal variability known to be present there³. A northward flow of $16 \pm 3 \text{ Sv}$ of thermocline water from the South Atlantic Ocean balances the NADW southward flow; the strength of the Pacific-Indonesian throughflow is estimated at $16 \pm 5 \text{ Sv}$, consistent with the most recent current-meter measurements⁸. This estimate is sensitive to

the use of hydrography at very low latitudes where noise susceptibility and seasonal variability are large.

Vertical exchanges (advection, denoted w^* , and mixing, measured by a diffusion coefficient κ^* across the neutral surfaces⁹ defining the layers) are associated with the general circulation. Both parameters are now the focus of intense research^{10,11} because numerical models are known to be very sensitive to κ^* . Direct measurements have shown little mixing in the ocean interior, while enhanced mixing of order $1,000 \text{ cm}^2 \text{ s}^{-1}$ has been observed in oceanic regions with rough topography¹¹. Table 1 summarizes the abyssal dianeutral (near vertical) exchanges by basin and depth range. Let an overbar denote a basin average. Then, in the deep range ($\approx 2,000$ – $3,500 \text{ m}$), dianeutral velocities are $\overline{w^*} = (0.1\text{--}0.6) \times 10^{-6} \text{ m s}^{-1}$, and the diffusivities are $\overline{\kappa^*} = (3\text{--}4) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$ with global averages of $\overline{w^*} = (0.13 \pm 0.03) \times 10^{-6} \text{ m s}^{-1}$ and $\overline{\kappa^*} = (3.7 \pm 0.7) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$ between 30°S and 47°N . In the bottom layers ($\approx 3,800 \text{ m}$ to the bottom), both means are larger: $\overline{w^*} = (0.4 \pm 0.1) \times 10^{-6} \text{ m s}^{-1}$ and $\overline{\kappa^*} = (9 \pm 2) \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$. Diffusivity values in the deep range are consistent with previous calculations from a crude one-dimensional global balance¹⁰. The largest values are found in the bottom layers although they are correspondingly more uncertain³. The spatial average values that are required by our tracer balance result from all mixing processes, including probable strong mixing generated near topography^{10,11}. We interpret those values¹⁰ as being the basin-average value of a process strongly mixing the ocean at highly localized regions.

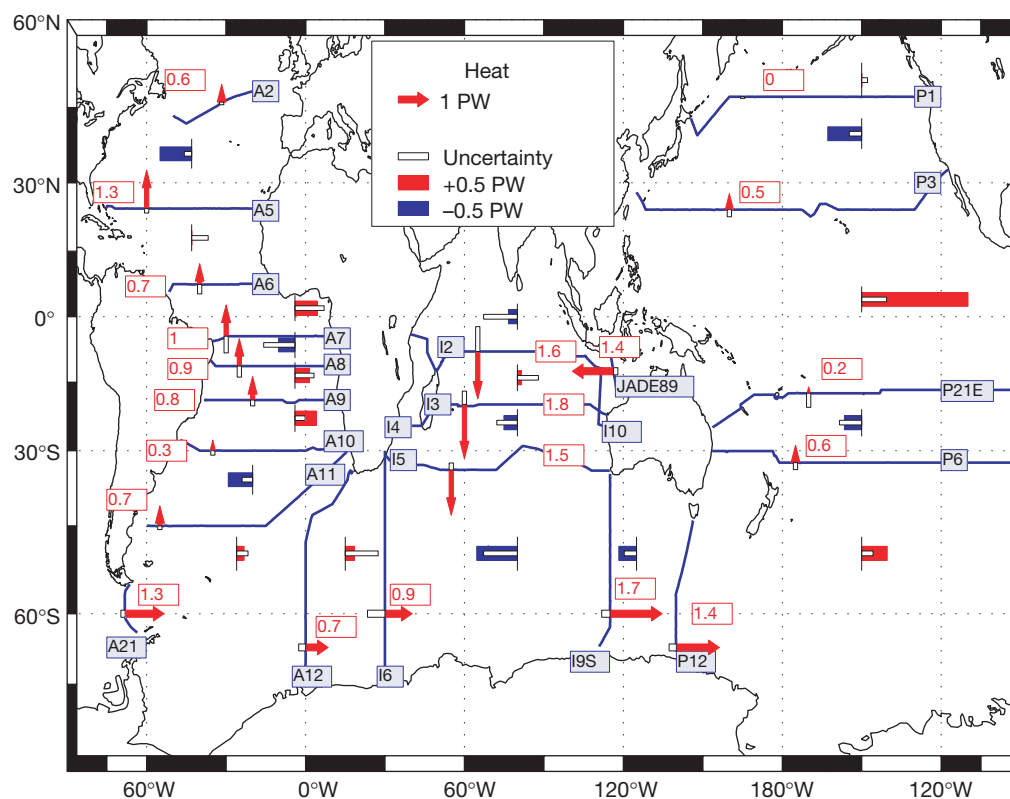


Figure 1 Hydrographic sections and heat fluxes. Transoceanic sections from the WOCE program were selected to ensure a reasonable temporal consistency (1990–1996) and to avoid crossing sections. Between Jakarta and Australia, the 1989 section from the Franco-Indonesian JADE program was used while pre-WOCE sections (1985, 1987)³ were used in the North Pacific and at 32°S in the Indian Ocean. Each section is a collection of high-density temperature, salinity, oxygen and nutrient measurements. From temperature and salinity, a geostrophic velocity field is calculated and adjusted so that mass and other conservative tracers (see Methods) are conserved between sections. The resulting heat (or 'enthalpy', where the net mass flux is non-zero) transports are indicated

by the arrows and red numbers (positive northward/eastward). The white box at the tail end of each arrow is the one-standard-deviation uncertainty. Between sections, ocean-atmosphere heat transfers are indicated by the zonal length of the coloured boxes (blue for ocean cooling; red for ocean heating), with the length of the white box inside indicating the uncertainty. (Because the ocean-atmosphere heat transfers are anomaly residuals, that is, corrected for residual mass imbalances, they do not correspond exactly to the differences between net fluxes across sections, for example, in the North Indian Ocean. But this discrepancy is much less than the uncertainties.)

Diffusivities could not be resolved in the Southern Ocean, where many neutral surfaces outcrop. The improved inverse model method has produced the first near-global, resolved estimates of the dianeutral transfers. The overall results are inconsistent with recent suggestions that the ocean mixes primarily at near-surface outcrops of the neutral surfaces, that is, primarily in the Southern Ocean¹². Strong abyssal mixing is required by the observed geostrophically balanced circulation, and its absence is incompatible with the observed property distributions.

Figure 1 shows the heat (actually, enthalpy) transports, across each hydrographic section (arrows) along with the residuals reflecting atmospheric heat exchanges (boxes). Residuals are accurately determined at middle and high latitudes, but are more uncertain at lower latitudes (for example, in the Atlantic Ocean) owing to an enhancement of the geostrophic noise there³. Nevertheless, the total heating over the tropical Atlantic and Pacific oceans are well-determined, respectively 0.7 ± 0.2 PW (1 PW = 10^{15} W) and 1.6 ± 0.4 PW. No significant heat transfers are found in the Indian Ocean because of the large, uncertain, warm water inflow from the Pacific Ocean. This large warm water flux is the main heat escape from the Pacific Ocean, resulting in a northward heat flux in the South Pacific. In the southern Pacific sector, significant heating is found, in contrast with the sparse *in situ* observations¹³, but in qualitative agreement with the recent re-analysis of the European Centre for Medium Range Weather Forecasts¹⁴. Figure 3 shows the globally integrated heat fluxes compared to independent estimates. Most of the cooling occurs in the Northern Hemisphere, at a rate of -1.7 ± 0.2 PW, in balance with the 2.3 ± 0.4 PW heating in the

tropical band and the -0.7 ± 0.3 PW cooling in the Southern Ocean.

Changes in the oceanic heat transport can have a large impact on atmospheric temperature gradients^{15,16} and thus on climate. Previous estimates of the ocean-atmosphere heat exchanges that are based upon purely ocean surface observations are highly uncertain^{17,18}. Analyses from numerical weather prediction centres provide oceanic surface fluxes that are often used as boundary conditions for driving ocean models, but associated uncertainty estimates are not provided. Heuristic calculations suggest uncertainties in their estimates of at least ± 0.6 PW for the meridional oceanic heat transport at most latitudes¹⁹. The present inversion indicates uncertainties that depend on latitude, with a high accuracy of globally integrated heat transfers (Fig. 3). Similar budgets, to be

Table 1 Basin-averaged dianeutral velocities and diffusivities

	$\bar{w}^*$ (10^{-6} m s $^{-1}$)	$\bar{\kappa}^*$ (10^{-4} m 2 s $^{-1}$)
Atlantic bottom	0.5 ± 0.2	9 ± 4
Indian bottom	0.6 ± 0.3	12 ± 7
Pacific bottom	0.4 ± 0.1	9 ± 2
Southern bottom	-0.25 ± 0.1	—
Atlantic deep	0.1 ± 0.05	3 ± 1.5
Indian deep	0.3 ± 0.15	4 ± 2
Pacific deep	0.1 ± 0.03	4 ± 1
Southern deep	0.1 ± 0.1	—

The average is calculated on neutral surfaces from $\gamma^* = 28.1$ kg m $^{-3}$ to the bottom (generally 3,800 decibars (or metres) to the bottom) for the 'bottom layers' and from $\gamma^* = 27.96$ to $\gamma^* = 28.07$ for the 'deep layers' (generally 2,000 m to 3,500 m).

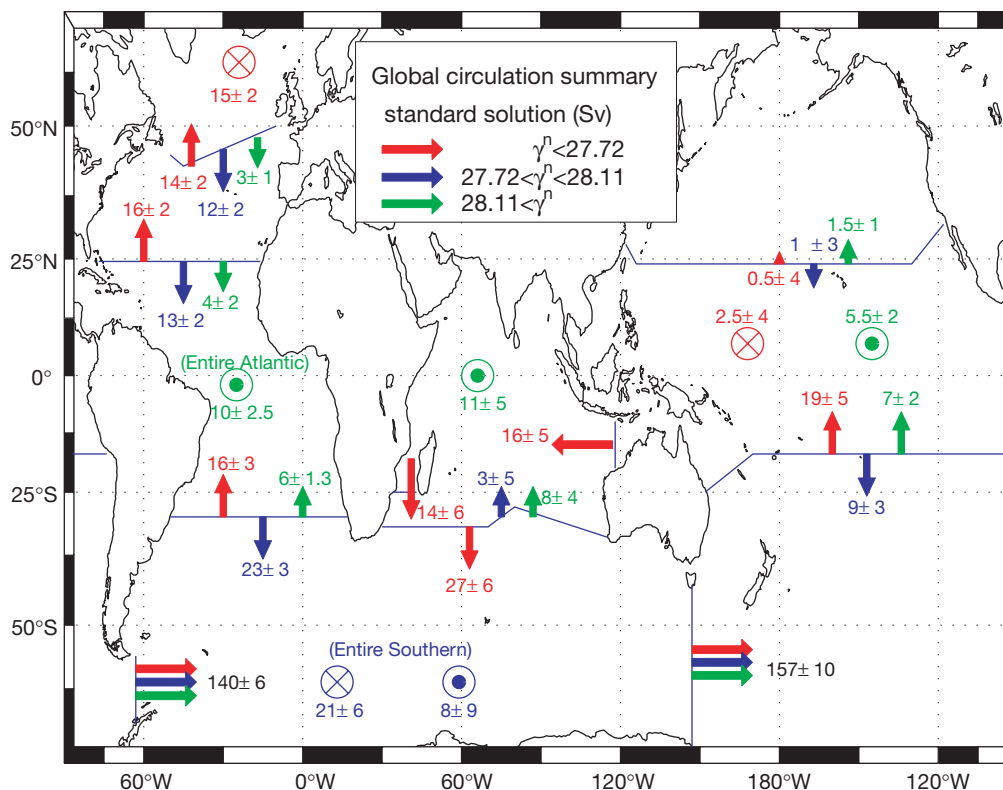


Figure 2 Zonally integrated layer mass transports. The estimated water transports are indicated for the different density classes bounded by neutral surfaces (γ^* , in kg m $^{-3}$) and across selected hydrographic sections. Neutral surfaces are close to surfaces of constant density, but are chosen so that movement along them minimizes the work done against gravity. The colour of the upwelling or down-welling arrows indicates the layer from which the water is coming. A flux of 0.8 Sv from the Pacific to the Atlantic Ocean through the Bering Strait was taken into account, although it is much smaller than the uncertainties on

the net transports. In the Southern Ocean, the bottom water formation takes place mostly in the Weddell Sea, while the upwelling distribution is uncertain. In the Indian Ocean, most of the upwelling takes place north of 7° S. The South Pacific transports are given at 17° S because of the more complicated structure at 32° S (ref. 3). Note the increase in the Southern Ocean transport south of Australia owing to the recirculation of Indonesian throughflow water.

published elsewhere, of oxygen and nutrients, are important for both climate and biogeochemical processes^{3,20}.

The physical circulation estimated here during WOCE provides a new reference state for climate change studies, and a test—within the estimated uncertainties—of general circulation models of both atmosphere and ocean. The estimated values of κ^* and w^* represent values that models will need to simulate. The high-resolution WOCE data set, the confinement in time, the use of current-meter data as constraints in boundary currents, and the detailed analysis of the uncertainties associated with our method are great improvements in reliability over previous circulation estimates. Within the error estimates, there is no conflict in either transports or heat fluxes with a previous global inversion that was based on a nearly independent and coarser data set^{1,21}, and no statistically significant change in integrated mass transports over the past 30 years was found^{1,22}. This is an important conclusion in its own right.

Improving the accuracy of the estimates made from the present data sets, if interpreted as climatological averages, will not be easy. Although the addition of new data, for example, existing or future meridional sections, will better the spatial resolution, limitations now lie primarily in the uncertainty introduced by true oceanic variability from the daily to the interannual. Significant improvements in the present numbers will occur only through the use of data sets permitting true temporal averaging of the oceanic circulation. For instance, the present solution has large uncertainties due to undersampling of the highly variable Brazil current and Pacific–Indian throughflow. Additional observations there would greatly improve accuracy. Further refinement of the a priori error estimates

by using improved numerical models and other data sets (such as altimetry) is also possible. □

Methods

The method is that of hydrographic inverse box models²³, where the relative, geostrophic velocity field is obtained from temperature and salinity measurements across the sections of Fig. 1. This initial flow is uniformly adjusted at each location so that the flow satisfies near-conservation of mass, anomalies of salt, heat, and the phosphate-oxygen combination (“PO” = $170[\text{PO}_4] + [\text{O}_2]$)^{24,25} within oceanic layers defined by neutral surfaces (neutral surfaces are densities chosen so that work against gravity is minimized when following the surface, permitting global use of a single density variable²⁶). Silica conservation is also required, top-to-bottom, while heat and PO are conserved only in layers that are not in contact with the surface. The surface layer is directly wind-driven (Ekman transport). Diffusive and advective exchanges between layers are determined by the model. The use of anomaly equations²⁷ permits, through noise subtraction, determination of diffusivities across each interface. Substantial improvements were made to the method, compared to that used in ref. 1, with, in particular, rigorous estimates of uncertainties through a determination of the a priori model error based on a simulation from the output of a quasi eddy-resolving ($1/4^\circ$) ocean general circulation model²⁸.

Received 24 February; accepted 4 September 2000.

- Macdonald, A. M. & Wunsch, C. An estimate of global ocean circulation and heat fluxes. *Nature* **382**, 436–439 (1996).
- Houghton, J. T. et al. (eds) *Climate Change 1995* (Cambridge Univ. Press, 1996).
- Ganachaud, A. *Large Scale Oceanic Circulation and Fluxes of Freshwater Heat, Nutrients and Oxygen*. Thesis, Massachusetts Institute of Technology, Woods Hole Oceanographic Institution, (1999).
- Ganachaud, A., Wunsch, C., Marotzke, J. & Toole, J. The meridional overturning and large-scale circulation of the Indian Ocean. *J. Geophys. Res.* (in the press).
- Roemmich, D. & McCallister, T. Large scale circulation of the North Pacific Ocean. *Prog. Oceanogr.* **22**, 171–204 (1989).
- Broecker, W. S., Sutherland, S. & Peng, T.-H. A possible 20th-century slowdown of Southern Ocean deep water formation. *Science* **286**, 1132–1135 (1999).
- Orsi, A. H., Johnson, G. C. & Bullister, J. L. Circulation, mixing and production of Antarctic Bottom Water. *Prog. Oceanogr.* **43**, 55–109 (1999).
- Gordon, A. L., Susanto, R. D. & Ffield, A. Throughflow within the Makassar Strait. *Geophys. Res. Lett.* **26**, 3325–3328 (1999).
- Jackett, D. R. & McDougall, T. J. A neutral density variable for the world's oceans. *J. Phys. Oceanogr.* **27**, 237–263 (1997).
- Munk, W. & Wunsch, C. The Moon and mixing: abyssal recipes II. *Deep-Sea Res.* **45**, 1977–2010 (1998).
- Polzin, K. L., Toole, J. M., Ledwell, G. R. & Schmitt, R. W. Spatial variability of turbulent mixing in the abyssal ocean. *Science* **276**, 93–96 (1997).
- Toggweiler, J. R. & Samuels, B. On the ocean's large-scale circulation near the limit of no vertical mixing. *J. Phys. Oceanogr.* **28**, 1832–1852 (1998).
- da Silva, A., Young, C. & Levitus, S. *Atlas of Surface Marine Data Vol. 1, Algorithms and Procedures* (NOAA Atlas NESDIS 6, US Dept of Commerce, 1994).
- Garnier, E., Barnier, B., Siefert, L. & Béranger, K. Investigating the 15-year air-sea flux climatology from the ECMWF reanalysis project as a surface boundary condition for ocean models. *Int. J. Climatol.* (in the press).
- Manabe, S. & Stouffer, R. J. Two stable equilibria of a coupled ocean-atmosphere model. *J. Clim.* **1**, 841–866 (1988).
- Marotzke, J. in *Decadal Climate Variability, Dynamics and Predictability* (eds Anderson, D. L. T. & Willebrand, J.) Vol. 1, 44 (NATO ASI/Springer, 1996).
- Gleckler, P. J. & Weare, B. C. Uncertainties in global ocean surface heat flux climatologies derived from ship observations. *J. Clim.* **10**, 2764–2781 (1997).
- Josey, S. A., Kent, E. C. & Taylor, P. K. New insights into the ocean heat budget closure problem from analysis of the SOC air-sea flux climatology. *J. Clim.* **9**, 2856–2880 (1999).
- Keith, D. W. Meridional energy transport: uncertainty in zonal means. *Tellus* **47**, 30–44 (1995).
- Ganachaud, A. & Wunsch, C. Oceanic nutrient and oxygen fluxes during the World Ocean Circulation Experiment and bounds on export production. *Glob. Biogeochem. Cycles* (submitted).
- Macdonald, A. The global ocean circulation: a hydrographic estimate and regional analysis. *Prog. Oceanogr.* **41**, 281–382 (1998).
- Schmitz, W. J. Jr & McCartney, M. S. On the North Atlantic circulation. *Rev. Geophys.* **31**, 29–49 (1993).
- Wunsch, C. *The Ocean Circulation Inverse Problem* 437 (Cambridge Univ. Press, 1996).
- Anderson, L. A. & Sarmiento, J. L. Redfield ratios of remineralization determined by nutrient data analysis. *Glob. Biogeochem. Cycles* **8**, 65–85 (1994).
- Broecker, W. S. “NO”, a conservative water-mass tracer. *Earth Planet. Sci. Lett.* **23**, 100–107 (1974).
- McDougall, T. Neutral surfaces. *J. Phys. Oceanogr.* **17**, 1950–1964 (1987).
- McDougall, T. J. in *Parameterization of Small-Scale Processes, Hawaiian Winter Workshop Aha Hulikoa* 355–386 (Univ. Hawaii, Manoa, 1991).
- Stammer, D., Tokmakian, R., Semtner, A. & Wunsch, C. How well does a $1/4^\circ$ global circulation model simulate large scale oceanic observations? *J. Geophys. Res.* **101**, 25779–25811 (1996).
- Trenberth, K. E. & Solomon, A. The global heat balance; heat transports in the atmosphere and ocean. *Clim. Dyn.* **10**, 107–134 (1994).
- de las Heras, M. & Schlitzer, R. On the importance of intermediate water flows for the global ocean overturning. *J. Geophys. Res.* **104**, 15515–15536 (1999).
- Bryden, H. L., Roemmich, D. H. & Church, J. A. Ocean heat transport across 24°N in the Pacific. *Deep-Sea Res.* **38**, 297–324 (1991).

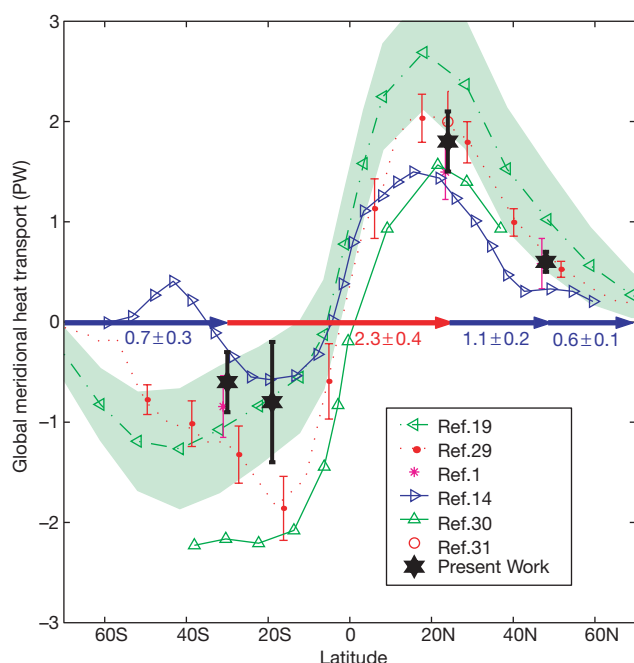


Figure 3 Global meridional heat transports. The black stars with thick error bars indicate the global, zonally integrated heat transports. Ocean–atmosphere heat exchanges between selected latitudes are given by numbers above the northward pointing arrows (positive for oceanic heating). The uncertainties from ref. 19 are indicated by the green band. The previous estimate of ref. 1 is different, but statistically consistent. No incompatibility is found between the estimates by the European Centre for Medium Range Weather Forecasts and US National Center for Environmental Prediction models^{14,19,29} and our new values, owing to their large uncertainties¹⁹. Other climatological estimates^{30,31} were published without error estimates, thus precluding any quantitative comparison. The estimate of ref. 30 is taken from their experiment B. (This figure is adapted and completed from refs 1 and 30.)

Acknowledgements

This work was completed while A.G. was a graduate student in the MIT/WHOI Joint Program in Physical Oceanography. We thank J. Toole, B. Warren, J. Marotzke, D. Glover and N. Hogg. B. Arbic, G. McKinley, A. Czaja, A. Macdonald, J. Marshall and M. Fieux also provided helpful comments on the manuscript. We are grateful to the principal investigators who provided the data from the World Ocean Circulation Experiment and the Franco-Indonesian Java-Australia Dynamic Experiment. G. Brown and D. Spiegel helped to design the figures. We were supported by the Jet Propulsion Laboratory and by gifts from Ford, General Motors and Daimler-Chrysler to MIT's Climate Modelling Initiative. This work is a contribution to the World Ocean Circulation Experiment.

Correspondence and requests for materials should be addressed to A.G. (e-mail: ganacho@ifremer.fr).

The use of earthquake rate changes as a stress meter at Kilauea volcano

James Dieterich*, Valérie Cayol†* & Paul Okubo‡

* US Geological Survey, Menlo Park, California 94025, USA

† Université B. Pascal, Clermont Ferrand, France

‡ Hawaiian Volcano Observatory, US Geological Survey, Hawaii 96718, USA

Stress changes in the Earth's crust are generally estimated from model calculations that use near-surface deformation as an observational constraint. But the widespread correlation of changes of earthquake activity with stress^{1–5} has led to suggestions that stress changes might be calculated from earthquake occurrence rates obtained from seismicity catalogues. Although this possibility has considerable appeal, because seismicity data are routinely collected and have good spatial and temporal resolution, the method has not yet proven successful, owing to the non-linearity of earthquake rate changes with respect to both stress and time. Here, however, we present two methods for inverting earthquake rate data to infer stress changes, using a formulation for the stress- and time-dependence of earthquake rates⁶. Application of these methods at Kilauea volcano, in Hawaii, yields good agreement with independent estimates, indicating that earthquake rates can provide a practical remote-sensing stress meter.

The inversions use a formulation for earthquake rate changes⁶ derived from laboratory observations of rate- and state-dependent fault strength^{6–8}, which constrain the earthquake nucleation process to be dependent on both time and stress. Previously, this formulation has been applied to model the spatial and temporal characteristics of earthquake clustering phenomena, including foreshocks and aftershocks^{6,7}, and to evaluate earthquake probabilities following large earthquakes⁹. The effectiveness of the formulation for forward modelling of earthquake phenomena, and its derivation from observed fault properties, provide the basis for its use to estimate stress changes from earthquake rate data. This approach yields stresses that drive the earthquake process. As such, it is distinct from other seismological methods that yield measures of stress changes resulting from earthquakes.

The formulation of Dieterich⁶ for rate of earthquake activity R (in a specified magnitude range) can be written in the condensed form

$$R = \frac{r}{\gamma \dot{S}_r}, \text{ where } d\gamma = \frac{1}{A\sigma} [dt - \gamma dS] \quad (1)$$

where γ is a state variable, t is time, and S is a modified Coulomb stress function defined below. The constant r is the steady-state earthquake rate at the reference stressing rate $\dot{S}_r$. A is a dimensionless fault constitutive parameter with values usually in the range 0.005–0.015 (refs 6–8). The modified Coulomb stress function is defined as

$$S = \tau - [\mu - \alpha]\sigma \quad (2)$$

where τ is the shear stress acting across fault planes that generate earthquakes (positive in the slip direction), σ is the normal stress (less pore fluid pressure), μ is the coefficient of fault friction and α is a constitutive parameter^{6,10} with an assigned value in this study of 0.25 (refs 6, 10). In equation (1), the term $A\sigma$ is a constant (that is, changes in σ are negligible relative to total σ). For a stress step, equation (1) yields the characteristic aftershock sequence, which consists of an immediate jump of seismicity rate followed by decay that obeys the Omori t^{-1} aftershock decay law with aftershock duration $t_a = A\sigma/\dot{S}_r$ (ref. 6).

We use two methods to estimate stress changes from earthquake rate data. The first gives stress as a function of time in a specified volume. From equation (1), the observed rate R is used to directly calculate γ as a function of time (that is, $\gamma(t) = r/R(t)\dot{S}_r$). This requires an estimate of $\dot{S}_r$, which can be obtained from independent

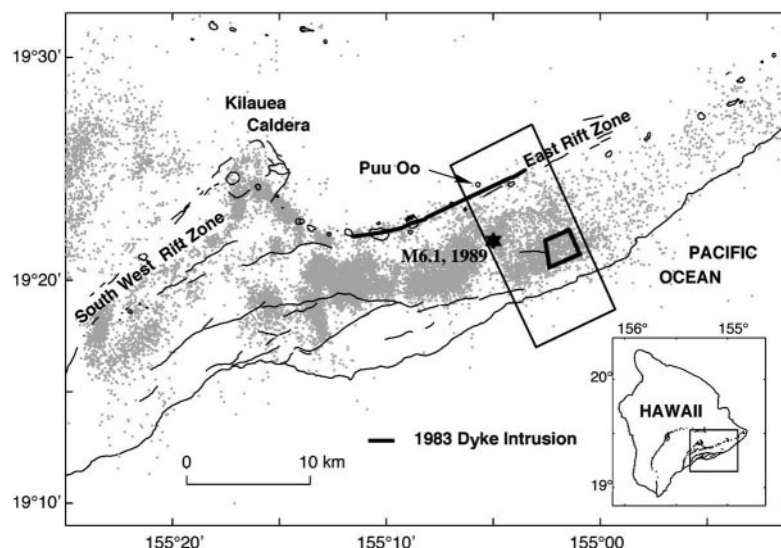


Figure 1 Map of Kilauea volcano showing earthquakes of magnitude $M \approx 1.5$ from 1976 to 1983. Intrusion of magma into the southwest and east rift zones of Kilauea causes rift expansion, and results in motion of the south flank to the SSE. Eruptive fissures initiated

the Puu Oo eruption, which started 1 January 1983 and continues to the present. The small polygon is the region of analysis of Fig. 2; the large rectangle gives the region of analysis of Figs 3 and 4.

observations of deformation rates, or from observations of after-shock decay using the relation given above for aftershock duration. Stress changes over successive time intervals are then obtained using the solution of equation (1) for a stress step at the mid-point of a time interval

$$\Delta S = A\sigma \ln \left[\frac{\gamma_i + \frac{\Delta t}{2A\sigma}}{\gamma_{i+1} - \frac{\Delta t}{2A\sigma}} \right] \quad (3)$$

where γ_i and γ_{i+1} are the estimated values of γ at the beginning and end, respectively, of the time step Δt .

The second method gives the spatial distribution of stress changes for a stress event, such as an earthquake mainshock or a magmatic intrusion. This method uses the solution of equation (1) for a history that consists of constant stressing at $\dot{S}_r$, followed by a stress step ΔS (the stress event), followed by constant stressing at $\dot{S}_r$. Solving equation (1) for this stressing history, and integrating the rates to obtain the expected number of events in time intervals, gives

$$\Delta S = A\sigma \ln \left\{ \frac{\dot{S}_r [\exp(N_2 \dot{S}_r t_2 / N_1 A\sigma) - 1]}{\dot{S}_r [\exp(\dot{S}_r t_2 / A\sigma) - 1]} \right\} \quad (4)$$

where N_1 is the count of earthquakes in time interval t_1 immediately before the stress event, and N_2 is the earthquake count in interval t_2 immediately following the event. As is usual in Coulomb stress

analysis, ΔS depends on fault orientation and slip direction. Consequently, in solving equation (4) for stress, counts of earthquakes (sorted by fault orientation) are made for sub-regions centred on grid points.

The south flank of Kilauea volcano, Hawaii, was chosen for this initial application of these methods for two reasons: first, this region experiences frequent stressing events from well studied causes, and second, various deformation observations have been made that can provide independent estimates of stress changes. This region is undergoing rapid deformation due to expansion of the active rift zones, and has high rates of seismic activity (Fig. 1). Since 1983, measured rates of south flank displacement^{11,12}, to the SSE, have ranged between 6 and 20 cm yr⁻¹. Expansion of the rift zones occurs by two mechanisms: (1) continuous slow intrusion of magma into a deep dyke-like body^{13,14}; and (2) rapid intrusion events that emplace shallow dykes, which sometimes form eruptive fissures^{11-13,15,16}. Rift zone expansion is coupled to, and accommodated by, fault slip at the interface between the volcano and the pre-existing sea floor¹⁶⁻¹⁸. The magnitude 6.1 earthquake of 25 June 1989, and numerous moderate and small earthquakes, occurred within the region of this study¹⁸⁻²⁰.

We have examined stresses in several sub-regions of the south flank of Kilauea using the method based on equation (3). The area indicated by the small polygon in Fig. 1 displays examples of seismicity and stress changes—shown in Fig. 2a and b, respec-

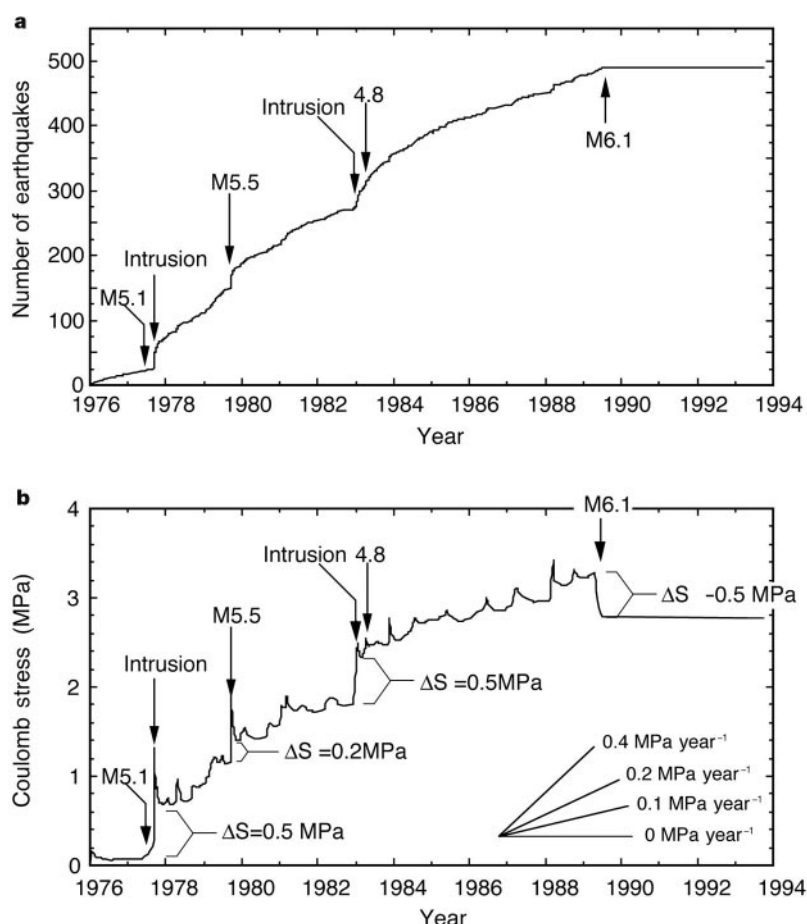


Figure 2 Earthquakes and stresses for the small polygon shown in Fig. 1. **a**, Cumulative number of earthquakes of magnitude $M \approx 1.5$ in the depth range 5–13 km. Transient changes of earthquake rates followed intrusion events of 1977 and 1983, the 1979 magnitude 5.5 earthquake, and the 1989 magnitude 6.1 earthquake. **b**, Stresses obtained from earthquake rates. A low pass filter was used to smooth short-term fluctuations in rate before calculating the stresses using equation (3) in the text. We

assume $A = 0.005$, σ equal to the overburden pressure less hydrostatic pore-fluid pressure using a bulk density of $\rho = 2.3$ for the rubbly lava flows of the volcano flank. These give a nominal value for $A\sigma = 0.45$ MPa at the representative earthquake depth of 7 km. The inversion employs $\dot{S}_r = 0.3$ MPa yr⁻¹; this value was obtained from the observed decay of earthquakes following the 1977 intrusion using the relation $t_a = A\sigma/\dot{S}_r$.

tively—that are representative of the region. At the times of the rift intrusion/eruption events of 12 September 1977 and 1 January 1983, and the magnitude 5.5 earthquake of 21 September 1979, seismic activity sharply increased and then decayed to a background rate following the t^{-1} Omori decay law. These seismicity changes

give Coulomb stress increases of about 0.5 MPa for the 1977 and 1983 intrusions. As discussed below, these stress changes are consistent with estimates from boundary element models constrained by geodetic data. The seismicity jump at the time of the 1979 earthquake gives a stress increase of 0.2–0.5 MPa. This is consistent with stresses obtained from elastic dislocation calculations for an earthquake of that magnitude at that distance (3–4.5 km distance, 2 km rupture radius, and 3 MPa stress drop). Aftershocks to the magnitude 6.1 earthquake of 25 June 1989 occurred on three sides of the polygon shown in Fig. 1, and gave stress increases of 0.3–0.6 MPa, but within the area shown, earthquake activity ceased following the 1989 mainshock. It is our interpretation that the polygon lies within a region of stress drop for the 1989 earthquake. The stress drop of 0.5 MPa in Fig. 2b is the minimum needed to reduce the rate to less than one event in the 5 years following the mainshock.

Various factors, including random fluctuations in earthquake rates and possible catalogue inconsistencies during swarm activity, can introduce artefacts into the inversions that can be confused with short-term stress changes. Consequently, short-term stress fluctuations shown in Fig. 2 should be treated with caution. However, long-term stressing rate changes and stress steps appear to be well resolved. The stress solution of Fig. 2b indicates a slowing of stressing rates, from about 0.30 MPa yr⁻¹ to about 0.15 MPa yr⁻¹, sometime between 1981 and 1983. This slowing is consistently seen for other polygons SSE of the 1983 event, and corresponds to a previously recognized¹¹ decrease of surface deformation rates over much of the south flank around the time of the 1983 intrusion, apparently because the continuing volcanic eruption decreased the supply of magma intruded into the rift.

Independent estimates of stress changes before the 1983 intrusion, and of stress changes resulting from it, are obtained from three-dimensional boundary element models^{20,21}; these stress changes are shown in Figs 3a and b, respectively. The models satisfy extensive deformation data sets before, and spanning, the 1983 intrusion; these data include horizontal displacements from trilateration measurements, level line data across the rift zone, and tide gauge data. In addition, the models incorporate topographic effects and the coupled interactions between the rift zone and the decollement fault at the base of the volcano. Before the 1983 intrusion event, earthquakes within the south flank concentrated in regions

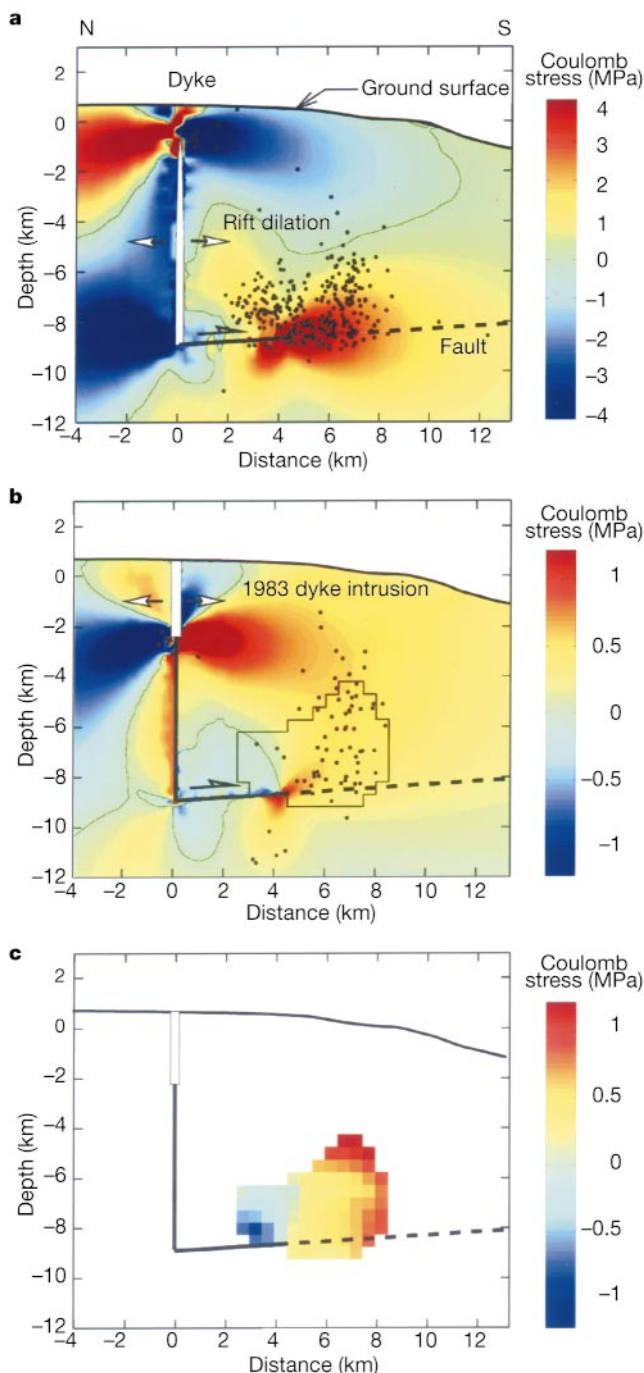


Figure 3 Earthquakes of magnitude $M \approx 1.5$ and Coulomb stress scale on a cross-section along the midline of the large box shown in Fig. 1. Coulomb stresses are calculated for faults that are parallel to the basal fault with slip perpendicular to the rift. Positive stress favours slip allowing rift expansion (arrows), and negative stress inhibits slip. The creeping section of the basal fault is shown as a solid line and the locked portion by a dashed line. **a**, Earthquakes and Coulomb stresses for the three years before the 1 January 1983 intrusion event and eruption²¹. **b**, Earthquakes for the 90 d following the 1 January 1983 eruption. Coulomb stress changes for the intrusion event are calculated for a pressurized dyke that extends to the surface to give the observed 2-m extension of trilateration lines that cross the rift zone. **c**, Stress changes obtained from the earthquake data using equation (4) in the text.

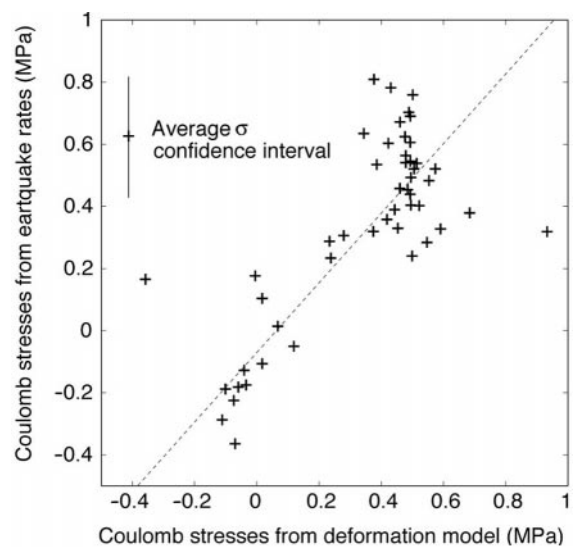


Figure 4 Comparison of stresses calculated from the boundary element model of the 1983 intrusion event (Fig. 3b) with the stresses calculated from the seismicity rate changes (Fig. 3c). Only the better-quality earthquake-derived stresses with $(N_1 + N_2) \approx 45$ (see text) are shown.

where stresses, resulting from slow inflation of the deep rift zone, are high (Fig. 3a). At the time of the shallow intrusion, the rift opened about 2 m (refs 11, 15, 21), an eruptive fissure formed (Fig. 1), and seismicity within this region of the south flank immediately increased by more than a factor of ten (Fig. 2a). The intrusion also resulted in changes in the patterns of stressing and seismic activity (Fig. 3a, b). Relative to the pre-intrusion period, seismic activity shifted away from the rift and more shallow events occurred. These changes are consistent with stresses that we have calculated for the 1983 intrusion (Fig. 3b). That is, the regions of increased and decreased earthquake rates correspond to regions of increased and decreased stress, respectively, resulting from the intrusion.

We believe that the boundary element model of the 1983 intrusion is also a reasonable representation of the stress changes for the 1977 intrusion. Each occurred in the same region^{15,20}, each resulted in very similar changes in seismicity patterns and rates, and the limited deformation data for the 1977 event are consistent with the more extensive 1983 data set.

Figure 3c gives the distribution of stress changes for the 1983 intrusion obtained from the seismicity data using equation (4). Counts of earthquakes for this calculation are made on a square grid with 0.5-km spacing of centres, over a radius of 1 km. A three-year period, t_1 , was used for counts before the event (N_1), and 90 days, t_2 , was used for counts following the event (N_2). We use the stressing rate results for the small polygon in Fig. 1 to set $\dot{\sigma}_i = 0.3 \text{ MPa yr}^{-1}$ before intrusion and $\dot{\sigma}_i = 0.15 \text{ MPa yr}^{-1}$ following the intrusion. Additionally, the parameter $A\sigma$ is made depth-dependent by allowing σ to increase with depth by the weight of the overburden less hydrostatic pore-fluid pressure.

For the region of the small polygon (Fig. 1), the boundary element model gives a stress increase of 0.3–0.6 MPa for the 1983 intrusion. This compares well with the seismicity-based stress change of 0.5 MPa obtained for both the 1977 and 1983 intrusion events using equation (3) (Fig. 2b). The pattern and magnitudes of stress changes at the time of the 1983 intrusion obtained from earthquake data using equation (4) (Fig. 3c) are in agreement with the boundary element calculation (Fig. 3b). Regression of the Coulomb stresses from the boundary element model of the intrusion against stresses from equation (4) has a slope of 1.1, and the correlation coefficient is 0.80 (Fig. 4).

In summary, seismicity changes coincide with documented deformation events for the south flank of Kilauea volcano. Inversions of earthquake rates for stress changes based on equation (1) give consistent results that agree with other estimates of stress. An additional internal consistency check of the earthquake rate formulation is provided by the previously described slowing of deformation rates that began around 1981–1983. As noted above, equation (1) predicts an inverse relationship between stressing rate and the time t_a for seismicity to decay to the background rate following a stress step. As the data of Fig. 2a show, the time for seismicity to return to a background rate following the 1983 intrusion is about twice as long as that following the 1977 intrusion; this is consistent with a halving of the stressing rate.

The methods presented here provide a way to use seismicity rate information in earthquake catalogues as a stress meter. This approach is able to resolve both positive and negative stress steps, as well as long-term changes in stressing rate. Because this approach does not depend on previous models of specific structures, it can also provide constraints on the models used to analyse observations of deformation. □

Received 22 March; accepted 24 August 2000.

1. Simpson, R. W. & Reasenber, P. A. Earthquake induced stress changes on central California faults. *US Geol. Surv. Prof. Pap.* **1550-F**, 55–89 (1994).
2. King, G. C. P., Stein, R. S. & Lin, J. Static stress changes and the triggering of earthquakes. *Bull. Seismol. Soc. Am.* **84**, 935–953 (1994).
3. Jaumé, S. C. & Sykes, L. R. Evolution of moderate seismicity in the San Francisco Bay region, 1850 to 1993: Seismicity changes related to the occurrence of large and great earthquakes. *J. Geophys. Res.* **101**, 765–789 (1996).

4. Harris, R. A., Simpson, R. W. & Reasenber, P. A. Influence of static stress changes on earthquake locations in southern California. *Nature* **375**, 221–224 (1995).
5. Stein, R. S. The role of stress transfer in earthquake occurrence. *Nature* **402**, 605–609 (1999).
6. Dieterich, J. H. A constitutive law for rate of earthquake production and its applications to earthquake clustering. *J. Geophys. Res.* **99**, 2601–2618 (1994).
7. Dieterich, J. H. & Kilgore, B. Implications of fault constitutive properties for earthquake prediction. *Proc. Natl Acad. Sci. USA* **93**, 3787–3794 (1996).
8. Scholz, C. H. Earthquakes and friction laws. *Nature* **391**, 37–42 (1998).
9. Stein, R. S., Barka, A. A. & Dieterich, J. H. Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophys. J. Int.* **128**, 594–604 (1997).
10. Linker, M. F. & Dieterich, J. H. Effects of variable normal stress on rock friction: observations and constitutive equations. *J. Geophys. Res.* **97**, 4923–4940 (1992).
11. Delaney, P. T., Miklius, A., Arnadottir, T., Okamura, A. & Sako, M. Motions of Kilauea volcano during sustained eruption from the Puu Oo and Kapaianaha vents, 1983–1991. *J. Geophys. Res.* **98**, 17801–17820 (1993).
12. Owen, S. P. et al. Rapid deformation of the south flank of Kilauea volcano, Hawaii. *Science* **267**, 1328–1332 (1995).
13. Delaney, P. T., Fiske, R. S., Miklius, Okamura, A. & Sako, M. Deep magma body beneath the summit and rift zones of Kilauea volcano. *Science* **247**, 1311–1316 (1990).
14. Gillard, D., Rubin, A. M. & Okubo, P. Highly concentrated seismicity caused by deformation of Kilauea's deep magma system. *Nature* **384**, 343–346 (1996).
15. Dvorak, J. J. et al. Mechanical response of the south flank of Kilauea volcano, Hawaii, to intrusive events along the rift zones. *Tectonophysics* **124**, 193–209 (1986).
16. Dieterich, J. H. Growth and persistence of Hawaiian volcanic rift zones. *J. Geophys. Res.* **93**, 4258–4270 (1988).
17. Ando, M. The Hawaiian earthquake of November 29, 1975: Low dip angle faulting due to forceful injection of magma. *J. Geophys. Res.* **84**, 7616–7626 (1979).
18. Got, J., Fréchet, J. & Klein, F. W. Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea. *J. Geophys. Res.* **99**, 15375–15386 (1994).
19. Klein, F. W., Koyanagi, R. Y., Nakata, J. S. & Tanigawa, W. R. The seismicity of Kilauea's magma system. *US Geol. Surv. Prof. Pap.* **1350**, 1019–1185 (1987).
20. Cayol, V. & Cornet, F. H. 3D mixed boundary elements for elastostatic deformation field analysis. *Int. J. Rock Mech. Min. Sci.* **34**, 275–287 (1997).
21. Cayol, V., Dieterich, J. H., Okamura, A. T. & Miklius, A. High rates of deformation prior to the 1983 Eruption of Kilauea Volcano, Hawaii. *Science* **288**, 2343–2346 (2000).

Acknowledgements

We thank A. Rubin for useful suggestions for this manuscript.

Correspondence and requests for materials should be addressed to J.D. (e-mail: jd Dieterich@usgs.gov).

Concurrent density dependence and independence in populations of arctic ground squirrels

Tim J. Karels & Rudy Boonstra

Division of Life Sciences, University of Toronto at Scarborough, 1265 Military Trail, Scarborough, Ontario, M1C 1A4, Canada

No population increases without limit. The processes that prevent this can operate in either a density-dependent way (acting with increasing severity to increase mortality rates or decrease reproductive rates as density increases), a density-independent way, or in both ways simultaneously^{1–3}. However, ecologists disagree for two main reasons about the relative roles and influences that density-dependent and density-independent processes have in determining population size^{4,5}. First, empirical studies showing both processes operating simultaneously are rare⁶. Second, time-series analyses of long-term census data sometimes overestimate dependence^{7,8}. By using a density-perturbation experiment^{9–12} on arctic ground squirrels, we show concurrent density-dependent and density-independent declines in weaning rates, followed by density-dependent declines in overwinter survival during hibernation. These two processes result in strong, density-dependent convergence of experimentally increased populations to those of control populations that had been at low, stable levels.

In the boreal forest of the southwestern Yukon, populations of arctic ground squirrels (*Spermophilus parryii plesius*) fluctuate by a factor of about three ($0.7\text{--}2.2\text{ ha}^{-1}$) in synchrony with the 10-year cycle of snowshoe hares¹³. These squirrel populations are sparser and more dynamic than those in open tundra and meadow habitats, where densities are higher— 5.5 ha^{-1} (ref. 14) to 16 ha^{-1} (ref. 15)—and apparently stable^{14,16}. Whereas squirrel populations in alpine and arctic habitats are thought to be limited by food availability, suitable burrowing habitat, and spacing behaviour^{14,16,17}, a large-scale 10-year experimental manipulation of food and predators by the Kluane boreal forest ecosystem project¹⁸ showed that squirrel populations in the boreal forest were limited by an interaction between food and predators^{19,20} that acted primarily through changes in reproduction²⁰. By design, the Kluane project examined limitation (processes that changed the average or equilibrium population density¹⁰), and was unsuitable for determining population processes that operate in a density-dependent manner.

At the completion of the Kluane project in spring 1996, densities of ground squirrels in four control areas averaged $1.6 \pm 0.15\text{ ha}^{-1}$; exclusion of predators increased the density of one population to 3.3 ha^{-1} ; food addition increased the density of squirrels in two other populations to 5.9 and 11.6 ha^{-1} ; and predator exclusion combined with food addition increased the density of one population to 30.1 ha^{-1} (ref. 20). After the predator-exclusion fences were removed and food addition ended, it took two years for densities of ground squirrels in the experimental sites to converge to those on the control sites (Fig. 1) in a strong density-dependent manner (Fig. 2). The strength of density-dependent population growth rate may be overestimated by regressing rates of population growth on population density⁸ when sampling variance is ignored, particularly in time-series analyses. However, variance in estimates of population size (Fig. 1) were nominal owing to the high trappability of arctic ground squirrels (77–95%)¹⁹. In contrast to the time-series approach, our experiment avoids problems associated with: (1) temporal autocorrelation of sequential data; (2) insufficiencies of annual data in detecting density-dependent processes that act on timescales shorter than a year (ref. 21); and (3) lack of ability to detect the mechanisms of density dependence.

To identify the mechanisms determining density dependence in

ground squirrel populations, we monitored reproduction (presence of lactation, weaning success and litter size), emigration and survival (active season survival by radio-telemetry and winter hibernation survival from trapping records) for two years after the cessation of the experimental treatments in all populations (see Methods). We also continued to supplement the food of two high-density populations (Food A, B) (Fig. 1). Two variables were strongly and significantly density dependent (Table 1): the proportion of females weaning a litter (Fig. 3a), and the proportion of females surviving winter hibernation (Fig. 3b). Together, both variables explained 90% of the variation in the annual rate of population change ($n = 11$, $p < 0.001$) as experimental densities converged towards control densities. Partial correlations indicated that overwinter survival accounted for 79% of the variance in annual rates of population change after differences were adjusted for changes in weaning rate. When population changes were adjusted for differences in overwinter survival, weaning rate accounted for 23% of the variance in annual rates of population change. The proportion of females lactating was weakly density dependent, but both summer survival and litter size were density independent (Table 1). Emigration of adults did not contribute to the decline of the experimental populations as none of the females fitted with radio-collars (1996, $n = 79$; 1997, $n = 51$) emigrated despite the high densities and lack of food resources on some sites. Juvenile males disperse independently of population density and resources²². In contrast, juvenile females remain near the natal area under a wide range of population densities and only disperse from food-supplemented populations at summer densities of 60 ha^{-1} (ref. 22). Such situations never occurred after the treatments were removed because juveniles were unlikely to be weaned at densities above 12 ha^{-1} in the absence of supplemental food (Fig. 3a).

Food resources were the primary factor determining the proportion of females weaning a litter. High-density populations that continued to receive supplementary food (Food A, B) had high weaning rates (82–89%) compared with unfed high-density populations which displaced them above the density-dependent relationships of unsupplemented populations (Fig. 3a). However, increased

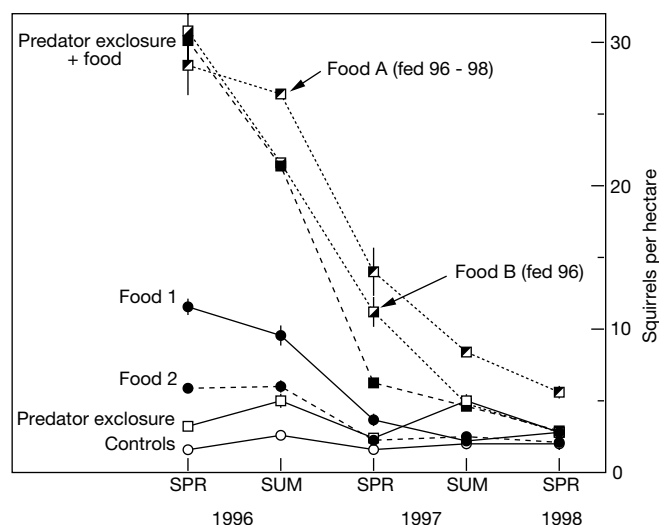


Figure 1 Convergence of arctic ground squirrel population densities after the removal of the experimental treatments of the Kluane project¹⁸ in spring 1996. Food addition (Food A, B) without predator exclusion slowed the rate of population decline, so that by 1998 spring densities (Food A) were still three times that of controls. Vertical error bars represent 1 s.e. of the population estimate for all experimental populations and 1 s.e. of the mean population estimate from four control populations. Points that appear without error bars have measures of error that are smaller than the symbol.

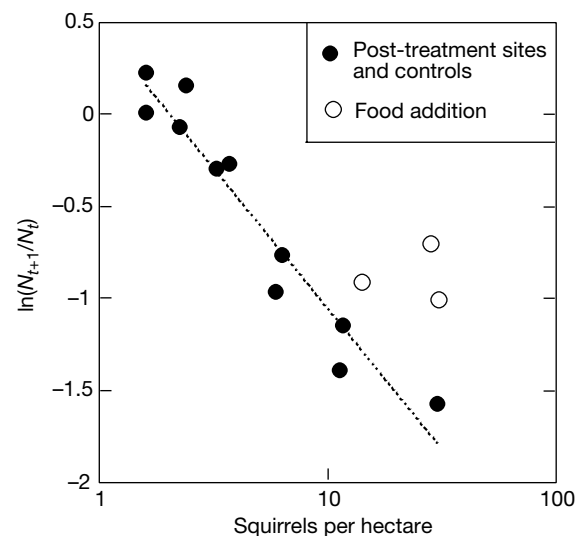


Figure 2 Density-dependent rates of population change of arctic ground squirrels after the experimental treatments of the Kluane project¹⁸ were removed in spring 1996. Here $\ln(N_{t+1}/N_t)$ is the rate of population change (N_{t+1} is the population size in the spring of year $t+1$; N_t is the population size in the spring of year t). Data are shown for controls and for the former experimental treatments. The equation of the dashed line is: rate of change = $-1.5[\log(\text{density})] + 0.5$, with $r^2 = 0.93$, $P < 0.001$. Data from sites where food was added were not included in the regression, but demonstrate that rates of increase were positively influenced by added food.

reproduction reduced, but did not prevent, the decline of the fed high-density populations (Fig. 1) compared with unfed high-density populations (Fig. 2). Fed high-density populations had higher overwinter survival (47–58%) than unfed high-density populations (11–20%), which again displaced them above the density-dependent relationships of unsupplemented populations (Fig. 3b). However, the food that we were supplying seemed to play a smaller role in determining overwinter survival than in determining weaning rate, because fed high-density populations had lower overwinter survival than unfed, low-density populations (60–95%) (Fig. 3b). The decline of the fed high-density populations was not accounted for by reduced summer survival as summer survival was very high (100% in 1996 and 86% in 1997), but was almost entirely attributable to poor overwinter survival. The food that we were providing was nutritionally optimized for the metabolic needs of rabbits, and may have lacked sufficient quantities of dietary components—possibly essential fatty acids—that are essential to hibernating mammals for reducing energy expenditure during hibernation²³. Together, the rabbit food that we supplied and the natural food was insufficient to maintain high overwinter survival in extremely dense ground squirrel populations.

Weaning rate showed strong density dependence, operating concurrently with density independence, when compared among years. Weaning rate maintained the same strength in density dependence over the two years (one-way analysis of covariance (ANCOVA): year $\times$ density, $F = 0.17$, d.f. = 1, $P = 0.90$), but females in 1997 were approximately 30% less likely to wean their litters independent of population density than were females in 1996 (one-way ANCOVA: $F = 57$, d.f. = 1, $P < 0.001$) (Fig. 3a). This indicates that changes in some regional environmental factor from 1996 to 1997 reduced weaning equally among all of the sites independently of population density, and offset the density-dependent relationship in 1997 below that in 1996. The density-independent effect was not caused by a change in food resources because the food addition site (Food A) also declined in weaning rate (7%) from 1996 to 1997, despite a 55% reduction in population density that should have increased weaning rate in 1997. Data from the Burwash Landing meteorological station (50 km from the study site) show that snow depth in the winter of 1995/96 was 36% greater than the 30-year average, whereas that in 1996/97 was about 30% lower. The greater snow depth in 1996 than 1997 should have maintained warmer soil temperatures during the winter²⁴ thereby reducing the energy expenditure of hibernating ground squirrels²⁵. Body condition of female squirrels, which is a strong predictor of reproductive success²⁶, was greater on all sites during the spring of 1996 than in spring 1997²⁷. We conclude that winter snow conditions acted as a

density-independent factor that altered reproduction in ground squirrels equally among all populations in the presence of strong density dependence.

The strength of density dependence in overwinter survival, as

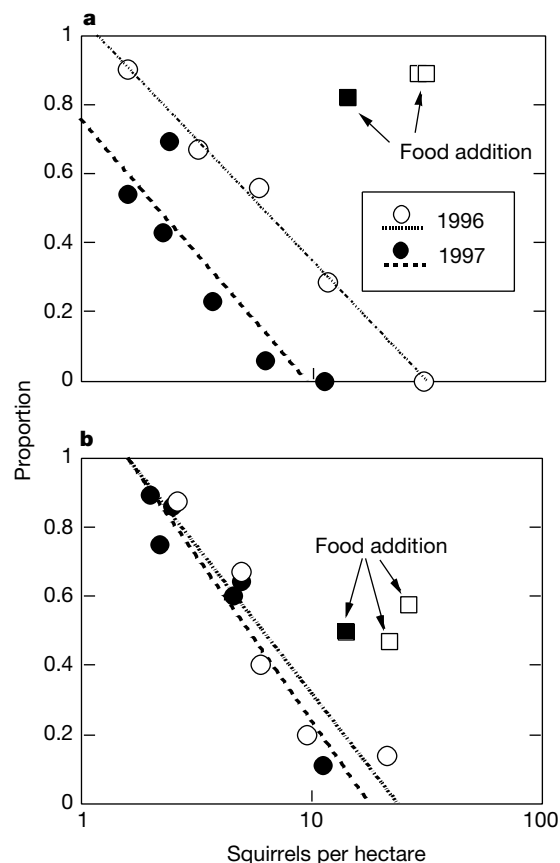


Figure 3 Mechanisms of population regulation in arctic ground squirrels. The figure shows density-dependence plots of the proportion of females weaning a litter (**a**) and the proportion of adult females surviving winter hibernation (**b**). Data are shown for controls and for former experimental treatments of the Kluane project¹⁸ after the treatments were removed during spring 1996. The density-dependent relationship in weaning rate in 1997 was offset from that in 1996 such that all squirrels in 1997 had a further 30% reduction in the probability of weaning a litter independent of population density; this demonstrates concurrent density dependence and independence.

Table 1 Regression statistics

Variable	Linear regression		Sample size of squirrels per site*									
	Year	Equation	r^2	N	P †	C	P-PE	P-F1	P-F2	P-PE+F	+FA	+FB‡
Lactating	1996	$-0.32x + 1.25$	0.70	5	0.076	22	24	44	25	80	47	39
	1997	$-0.88x + 1.50$	0.55	6	0.092	16	16	17	9	28	16	11
Weaning	1996	$-0.94x + 1.48$	0.97	5	< 0.003	10	9	9	7	15	9	9
	1997	$-1.11x + 1.16$	0.89	6	< 0.005	9	9	9	7	16	6	10
Litter size	1996	$-0.38x + 2.89$	0.19	4	0.57	9	13	6	6	0	28	15
	1997	$-1.57x + 3.21$	0.11	4	0.68	6	7	4	4	0	5	0
Summer survival	1996	$0.07x + 1.11$	0.01	5	0.87	10	9	11	11	15	11	12
	1997	$0.49x + 0.88$	0.06	5	0.70	10	8	8	10	8	7	0
Overwinter survival	1996	$-0.95x + 1.54$	0.88	5	0.019	21	34	53	31	131	32	36
	1997	$-1.08x + 1.08$	0.91	6	< 0.004	19	22	12	7	25	9	8

Reproduction and survival processes were analysed for density dependence by linear regression. All proportional data were arcsine transformed and densities were log-transformed before regression. Sample sizes given are the number of squirrels on each site that were used to calculate each variable before regression. Sites 'Food A' and 'Food B' were not included in the regression when food supplemented.

* Sites are abbreviated as follows: control (C), post-predator exclosure (P-PE), post-food 1 (P-F1), post-food 2 (P-F2), post-predator exclosure + food (P-PE+F), food addition A (+FA), and food addition B (+FB).

† Adjusted by the Bonferroni method for multiple tests of significance.

‡ not food-supplemented in 1997.

indicated by the regression coefficient, remained similar between years (one-way ANCOVA: year $\times$ density, $F = 0.251$, d.f. = 1, $P = 0.63$) as was the rate of overwinter survival when the effects of density were removed (one-way ANCOVA: $F = 0.032$, d.f. = 1, $P = 0.86$) (Fig. 3b). Therefore, overwinter survival maintained a constant relationship with density regardless of the changes in winter conditions that operated on weaning rate in a density-independent manner.

Though populations rarely, if ever, maintain or achieve an equilibrium density because of fluctuations in population growth induced by stochastic events^{3,10}, this does not imply that density-dependence is unimportant or does not exist. Processes that operate in a density-dependent manner return a population towards an equilibrium density and are regulating¹⁰. Experimental density manipulations are the strongest and most robust approach to teasing apart the relative roles of density-dependent and density-independent determination of population density^{9,11,12}. Our study demonstrates the power of this approach by showing how different density-dependent processes can operate to sequentially regulate population density in the presence of density-independent processes. Despite environmental differences between years that altered reproduction in a density-independent manner, density-dependent declines in reproduction remained constant between years and thus served as a dampening force against stochastic changes in reproduction. However, density-dependent weaning rate had a more limited role in regulating arctic ground squirrel populations than did a second and sequential density-dependent process, overwinter survival. During our study, density-dependent overwinter survival dominated density-independent changes in weaning rate resulting in strong density-dependent rates of population change—a necessary condition of population regulation². □

Methods

Spring and summer mark–recapture population estimates of arctic ground squirrels were conducted on areas of 8–10 ha from spring 1996 to spring 1998 within each of four control areas and within the following four former experimental areas of the Kluane project (1987–1996) described in detail elsewhere^{18,20}: two food-addition areas (36 ha), a predator-exclusion area (1 km²), and a food-addition area (36 ha) enclosed within a predator-exclusion area (1 km²). All fences were dismantled and food addition was discontinued in spring 1996 with the exception of two 2.5-ha areas (Food A, B) within the former predator-exclusion area with food-addition site where we continued weekly food supplementation by hand to all burrow sites during 1996; 20 kg of rabbit food per ha. Food B was not supplemented after 1996 but was monitored as ground squirrel densities on that site converged with those on the other sites. Population estimates and their standard errors were calculated using a closed population mark–recapture heterogeneity model (jack-knife) from program CAPTURE²⁸. Weaning success was calculated as the proportion of all females that had litters appearing above ground. Litter size was determined by intensive live-trapping at the natal burrow at the time when young appeared above ground. Dispersal and summer survival was based on radiotelemetry of adult females. Overwinter survival was calculated from trapping records as the proportion of individuals captured in spring that were present in the population before hibernation in the previous year. The details of our procedures for weaning success, litter size, dispersal and survival are described elsewhere^{20,22}.

Received 19 May; accepted 7 September 2000.

- May, R. M. Stability in randomly fluctuating versus deterministic environments. *Am. Nat.* **107**, 621–650 (1973).
- Turchin, P. in *Population Dynamics: New Approaches and Synthesis* (eds Cappuccino, N. & Price, P. W.) 19–40 (Academic, London, 1995).
- Sinclair, A. R. E. & Pech, R. P. Density dependence, stochasticity, compensation and predator regulation. *Oikos* **75**, 164–173 (1996).
- Strong, D. R. in *A New Ecology* (eds Price, P. W., Slobodchikoff, C. N. & Gaud, W. S.) 313–327 (Wiley, New York, 1984).
- Murray, B. J. Jr Can the population regulation controversy be buried and forgotten? *Oikos* **84**, 148–152 (1999).
- Leirs, H. et al. Stochastic seasonality and nonlinear density-dependent factors regulate population size in an African rodent. *Nature* **389**, 176–180 (1997).
- Ostfeld, R. S., Canham, C. D. & Pugh, S. R. Intrinsic density-dependent regulation of vole populations. *Nature* **366**, 259–261 (1993).
- Shenk, T. M., White, G. C. & Burnham, K. P. Sampling-variance effects on detecting density dependence from temporal trends in natural populations. *Ecol. Monogr.* **68**, 445–463 (1998).
- Murdoch, W. W. Population regulation and population inertia. *Ecology* **51**, 497–502 (1970).
- Sinclair, A. R. E. in *Ecological Concepts* (ed. Cherret, J. M.) 197–241 (Blackwell, Oxford, 1989).
- Harrison, S. & Cappuccino, N. in *Population Dynamics: New Approaches and Synthesis* (eds Cappuccino, N. & Price, P. W.) 131–147 (Academic, London, 1995).

- Cappuccino, N. & Harrison, S. in *Frontiers of Population Ecology* (eds Floyd, R. B., Sheppard, A. W. & De Barro, P. J.) 53–64 (CSIRO Publishing, Melbourne, 1996).
- Boutin, S. et al. Population changes of the vertebrate community during a snowshoe hare cycle in Canada's boreal forest. *Oikos* **74**, 69–80 (1995).
- Carl, E. A. Population control in arctic ground squirrels. *Ecology* **52**, 395–413 (1971).
- Lacey, E. A., Wiczorek, J. R. & Tucker, P. K. Male mating behaviour and patterns of sperm precedence in Arctic ground squirrels. *Anim. Behav.* **53**, 565–779 (1997).
- Batzli, G. O. & Sobaski, S. T. Distribution, abundance, and foraging patterns of ground squirrels near Atkasook, Alaska. *Arct. Alp. Res.* **12**, 501–510 (1980).
- Green, J. E. *Population Regulation and Annual Cycles of Activity and Dispersal in the Arctic Ground Squirrel*. Thesis, Univ. British Columbia (1977).
- Krebs, C. J. et al. Impact of food and predation on the snowshoe hare cycle. *Science* **269**, 1112–1115 (1995).
- Hubbs, A. H. & Boonstra, R. Population limitation in Arctic ground squirrels: effects of food and predation. *J. Anim. Ecol.* **66**, 527–541 (1997).
- Karels, T. J., Byrom, A. E., Boonstra, R. & Krebs, C. J. The interactive effects of food and predators on reproduction and overwinter survival of arctic ground squirrels. *J. Anim. Ecol.* **69**, 235–247 (2000).
- Leewellen, R. H. & Vessey, S. H. The effect of density dependence and weather on population size of a polyvoltine species. *Ecol. Monogr.* **68**, 571–594.
- Byrom, A. E. & Krebs, C. J. Natal dispersal of juvenile arctic ground squirrels in the boreal forest. *Can. J. Zool.* **77**, 1048–1059 (1999).
- Florant, G. L. Lipid metabolism in hibernators: the importance of essential fatty acids. *Am. Zool.* **38**, 331–340 (1998).
- Buck, C. L. & Barnes, B. M. Temperatures of hibernacula and changes in body composition of arctic ground squirrels over winter. *J. Mamm.* **80**, 1264–1276 (1999).
- Anufriev, A. I. & Akhremenko, A. K. Energetic cost of hibernation of the Arctic ground squirrel. *Sov. J. Ecol.* **21**, 279–283 (1990).
- Risch, T. S., Dobson, F. S. & Murie, J. O. Is mean litter size the most productive? a test in Columbian ground squirrels. *Ecology* **76**, 1643–1654 (1995).
- Karels, T. J. *Reproduction, Hibernation, and Population Regulation of Arctic Ground Squirrels (Spermophilus parryi plesius)*. Thesis, Univ. Toronto (2000).
- Pollock, K. H., Nichols, J. D., Browne, C. & Hines, J. E. Statistical inference for capture-recapture experiments. *Wildl. Monogr.* **107**, 1–97 (1990).

Acknowledgements

We thank M. Evans, E. Gillis, E. Gold, L. Hofer, J. Hoyle, H. Jewell, J. Lindell, C. McColl, C. Riley, W. Ruscoe, C. Walasek and J. Wilson for field assistance, and D. S. Hik, C. J. Krebs and J. R. Rosenfeld for comments and suggestions on the manuscript. This study was funded by a Northern Studies Training Program grant (Department of Indian and Northern Affairs) to T.J.K., a Natural Sciences and Engineering Research Council (NSERC) grant to R.B. and a NSERC collaborative special project grant to the Kluane boreal forest ecosystem project. T.J.K. was supported by a fellowship from the University of Toronto.

Correspondence and requests for materials should be addressed to R.B. (e-mail: boonstra@scar.utoronto.ca).

Variation in the reversibility of evolution

Henrique Teotónio*† & Michael R. Rose*

* Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92612, USA

How reversible is adaptive evolution^{1–6}? Studies of microbes give mixed answers to this question^{6–11}. Reverse evolution has been little studied in sexual populations^{12–14}, even though the population genetics of sexual populations may be quite different. In the present study, 25 diverged replicated populations of *Drosophila melanogaster* are returned to a common ancestral environment for 50 generations. Here we show that reverse evolution back to the ancestral state occurs, but is not universal, instead depending on previous evolutionary history and the character studied. Hybrid populations showed no greater tendency to undergo successful reverse evolution, suggesting that insufficient genetic variation was not the factor limiting reverse evolution. Adaptive reverse evolution is a contingent process which occurs with only 50 generations of sexual reproduction.

† Present address: University Museum of Zoology, Downing Street, Cambridge CB2 3EJ, UK.

Over the past 20 years, we have created an artificial evolutionary radiation by imposing demographic and stress selection on large, outbred, replicated populations of *Drosophila melanogaster*. Part of this radiation is shown in Fig. 1. The five different groups of replicated populations are multifold differentiated from each other for several life-history, physiological and biochemical characters^{15–21}. One of these groups has been maintained under the ancestral 'Ives' conditions since 1980 (the 'B' populations). One has been selected for reproduction late in life (the 'O' populations). One has been selected for survival under conditions of complete starvation (the 'SO' populations). One has been selected for reproduction in mid-life (the 'CO' populations). Finally, one group has been selected for reproduction very early in life (the 'ACO' populations) (see Fig. 1 legend and Methods). Most of the differentiation occurred within 50 generations of the start of selection.

At the start of the present study, a new population was derived from each of these 25 populations and then returned to the ancestral-Ives environmental conditions for 50 generations. The evolutionary response of these populations was assayed using eight characters which are related to fitness or were highly differentiated from the ancestral condition. By measuring these characters approximately every eight generations we were able to determine reverse evolutionary patterns and level of convergence.

The patterns of reverse evolution varied among evolutionary histories and among characters. Analysis of the evolutionary trajectories indicated that, except for fecundity, populations with different evolutionary histories had statistically heterogeneous responses to the reimposition of the ancestral environment (analysis of covariance, ANCOVA F -tests for evolutionary history and interaction term; $P < 0.05$). Results are shown in Fig. 2. Four kinds of evolutionary trajectories were observed. The first was reversion to ancestral character values with full convergence after little more than 20 generations. This was the case, for example, for IO and ISO female developmental time ('I' indicates a derived population; see Methods). The second kind of trajectory was a pattern of response without convergence within the 50 generations of the study, as in the case of IACO developmental time. The third kind of trajectory was

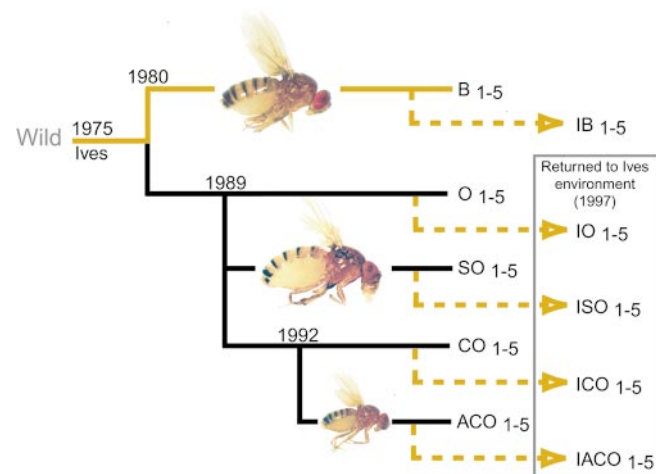


Figure 1 Laboratory radiation due to five selection treatments. Each group of populations was fivefold replicated (shown as subscript numbers). All populations descend from the Ives population, sampled from nature in 1975 (see Methods). We show female flies of the B, SO, and ACO treatments. These were cultured in identical conditions and are shown at the same relative scale. The direct descendants employed in our reverse evolution study are shown by dashed branches. Populations represented by yellow branches were maintained in the ancestral-Ives environment. Hybrid populations were derived either from crosses within or between different selection histories. A total of 44 populations were returned *de novo* to the ancestral-Ives environment and followed for fifty generations. Eight highly differentiated characters were measured during reverse evolution.

initial rapid reversion, followed by a stalling of evolution, without full convergence to ancestral character values, as for ISO starvation resistance or ICO developmental time. In this case, new stable states may have been attained. Finally, the fourth pattern was no significant change throughout the 50 generations of evolution, the best examples being ICO and IACO fecundity under high-density assay conditions.

The results show that reverse evolution back to an ancestral condition may or may not occur, at least over a 50-generation period of selection. Reverse evolution is neither inevitable nor impossible; it is contingent. This raises the question of the genetic mechanisms responsible for this contingency. Could populations that fail to return to the ancestral state lack genetic variation? Alternatively, could these populations be stuck in the vicinity of

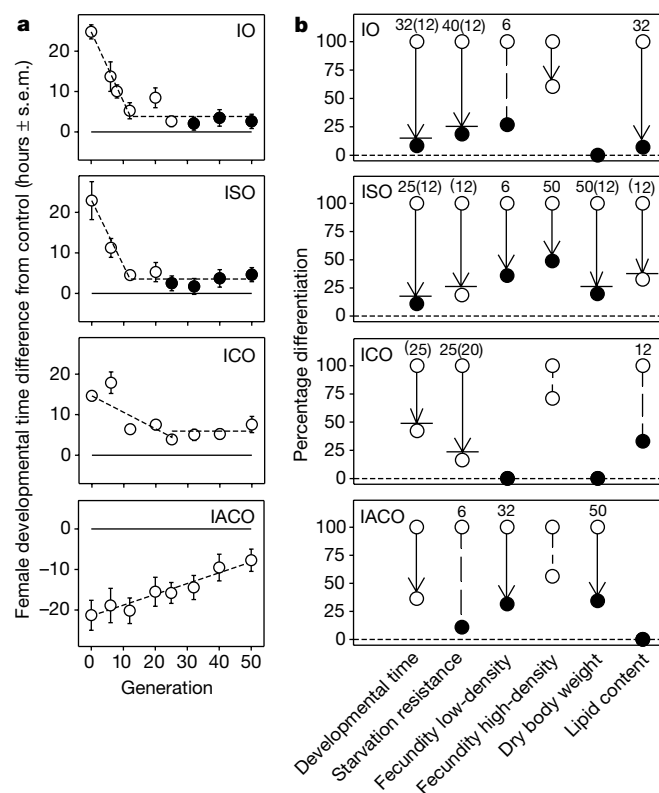


Figure 2 Reverse evolution during 50 generations. **a**, Response to selection in populations of different evolutionary history for female developmental time as a function of time (generations). Open circles show that the average difference from the control levels of the five replicate populations is significant (t -tests; $P < 0.05$). Black circles show non-significant differences from the controls, indicating convergence to ancestral levels. Dashed lines represent the best-fit regression models (IO, $P < 0.001$, adjusted $R^2 = 0.92$; ISO, $P < 0.001$, adjusted $R^2 = 0.95$; ICO, $P = 0.008$, adjusted $R^2 = 0.68$; IACO, $P < 0.001$, adjusted $R^2 = 0.93$). **b**, The categorized x-axis shows the response of each character separately, and the y-axis represents the percentage of differentiation relative to the control populations. Open circles indicate non-convergence to the control levels ($P < 0.05$), and black circles indicate convergence, as before. For all characters in all groups of populations, 100% differentiation was the initial starting condition, with the exception of a few characters which were undifferentiated from the beginning, these being represented by a single black circle at the zero differentiation line. Arrows between circles represent significant one-phase models, and arrow-lines indicate that a significant model with a first linear phase of convergence followed by a second no-response phase was the best fit to the evolutionary trajectory (see Methods). Broken lines show that no significant model was found, that is, that no significant change occurred during evolution. Numbers above the character for each group of populations show the generation of convergence to the control levels, and numbers in parentheses show the generation at which reverse evolution stalled. Reverse evolution patterns differ according to the character considered and according to selection history.

selective equilibria at which epistasis and linkage disequilibrium prevent their return to an earlier evolutionary state at which higher fitness would be attained? Both of these hypotheses can be tested by creating hybrids of the differentiated populations before reverse evolution. Hybrids should have more genetic variation (if the uncrossed experimental populations have become genetically impoverished and they should have undergone perturbation away from a strong evolutionary attractor (if that is their initial condition) through drastic gene frequency changes, subsequently coupled with recombination. If these genetic mechanisms are responsible for constraints on reverse evolution then we expect that hybrids should show more reverse evolution in the same 50 generations.

Nineteen hybrid populations were derived (see Methods). These underwent reverse evolution for 50 generations in parallel with the uncrossed derivatives already discussed. The appropriate test for improved convergence of the hybrids compared to the uncrossed must take into account the quantitative differentiation from the ancestral state in the two types of population. Many of the hybrids are less differentiated from the ancestral condition at the outset of reverse evolution, because of their hybridization. Therefore, we examined whether the hybrids and the uncrossed populations fall on the same regression of evolutionary rate on initial differentiation. Figure 3 shows this relationship among the 44 populations. There is no detectable difference between hybrid and uncrossed populations in their reverse evolution, once initial differentiation is allowed for. This indicates that the factors limiting reverse evolution are probably not related to a lack of genetic variation, epistasis, or linkage disequilibrium.

To account for these results we propose instead that the return to

the ancestral-Ives environment did not produce uniform selection pressures among populations of different evolutionary histories. This can be attributed to variation in the relationship of each character with fitness in the ancestral environment according to the specific genetic background of each experimental population. Selection in the environments imposed before reverse evolution could have created different genetic architectures among groups of populations, in turn producing different genetic relationships between these characters and fitness in the ancestral environment. For example, populations which take longer to develop than the controls respond faster to natural selection on developmental time than populations which began by taking less time to develop. Thus, in the former populations, developmental time is intimately connected with fitness in the ancestral environment, whereas in the latter, the relationship with fitness is weaker in the ancestral environment (see Fig. 2).

Despite the cases in which reverse evolution fails to occur, or is stalled, our study shows an overwhelming tendency to evolutionary reversal during 50 generations of selection for adaptation to the ancestral environment. This demonstrates the power of selection in sexual populations to achieve adaptation^{14,22–25}. History still plays a significant role in the observed evolutionary dynamics and outcomes^{3,4,26,27}, although this history is only a few hundred generations in duration. □

Methods

Experimental populations

We studied four different evolutionary histories, fivefold replicated, all derived from a common ancestor (Ives). These independent evolutionary groups included populations with 95 generations of selection history for increased late-life fertility (the O_{1-5})¹⁵, with 100 generations of selection for increased starvation resistance (the SO_{1-5})¹⁷, with 4–5-week intermediate generation times (the CO_{1-5})¹⁷, and 190 generations of selection for decreased developmental time (the ACO_{1-5})²¹. Derived populations were designated by 'I' followed by the acronym for their evolutionary history. The five populations which had always been maintained in the ancestral-Ives conditions served as the control populations: IB_{1-5} (Fig. 1). These have been maintained for approximately 440 generations¹⁵. At the start of the experiment, hybrids within each selection treatment were created by five-way crosses of all five replicate populations. One of these five-way crosses used the B populations, which served as our control hybrid populations after replicating the cross five times. Hybrids were also created by crosses among all possible combinations of different selection treatments. Hybrid and uncrossed populations were followed in parallel during the 50 generations of reverse evolution. The ancestral-Ives environmental conditions include discrete two-week generations in vials, with egg and adult density controlled, and egg laying taking place during about two hours after mixing all adults of the same population. The amount of nutrients, temperature, humidity and light schedule were controlled. Census population sizes were always above 1,000 individuals per replicate population.

Assay designs

Eight characters were analysed in this study, most measured under conditions close to those of the ancestral-Ives environment. Developmental time in both genders was measured as the time from egg to adult emergence²¹. A total of 600 eggs per population were measured in each assay. Starvation resistance in both males and females was measured as the time until death under starvation in high humidity conditions¹⁷. A total of 40 flies per gender and per population were measured in each assay. Early fecundity under high adult density conditions was measured as the number of eggs laid by 20 pairs of flies per vial during one hour. Sample sizes for this character were eight vials per population per assay. Early fecundity under low adult density conditions was measured as the number of eggs laid by a pair of flies per vial during 24 hours, the sample size being 40 such vials per population. Female dry body weight was calculated from the weight of 8 samples of 10 flies each per population, to the nearest 0.001 mg. Female whole-body lipid content was measured as the difference in dry weight before and after lipid extraction with a Soxhlet petroleum ether extractor.

Statistical analysis

All the starting populations employed in this study have been maintained independently of one another throughout their history and as such each one is an evolutionary experiment that is treated as a unique data point in our statistical analysis. Among the uncrossed populations, our experimental design allows a separation of genetic sampling effects from the main selection treatment effects. The IB_{1-5} are the control populations used to test the hypothesis of convergence to the ancestral character states. In each assay generation, each group of populations was independently compared for differences from the control populations by one-tailed unpaired Student *t*-tests. The evolutionary trajectories of the uncrossed populations were analysed with two general least-squares regression models to explain the response to selection: a one-phase linear model, with a constant

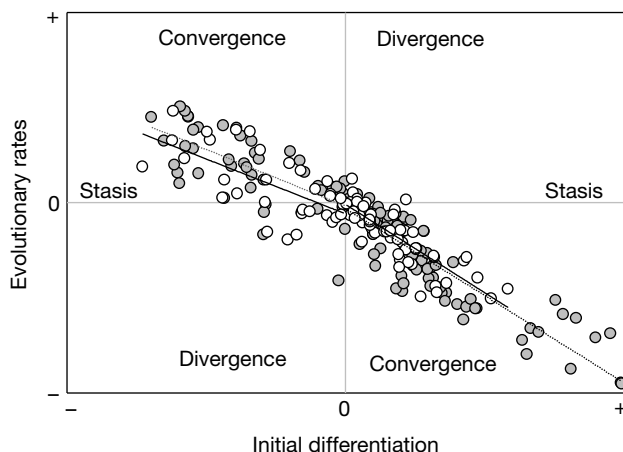


Figure 3 Reverse evolutionary rates as a function of differentiation at the start of the study. For each character and population, we estimated both the slope and the intercept of a one-phase linear regression, which were then used as the data points plotted, standardized to a common scale. Standardization involved dividing the estimated slopes and intercepts for each character by the largest slope or intercept for all data from that character, respectively. Data from fecundity under low adult density conditions were not used in this analysis. Uncrossed populations are shown by grey circles and hybrids by open circles. Out of 238 independent data points 209 fall in the two convergence quadrants, while only 29 fall in the two divergence quadrants. Assuming that stasis occurs when the evolutionary rate is less than 1% per generation, only 6 points fall in this class. The probability of chance occurrence of this pattern is much lower than 0.001 (sign test for convergence versus divergence). Hybrid and uncrossed evolutionary rate comparisons by ANCOVA (see Methods) did not reveal significant differences, when all characters were analysed either together or separately. Hybrids are not more likely to converge than uncrossed populations. For a significance level $\alpha = 0.05$, the statistical power to detect a difference between these groups, as estimated from the ANCOVA model, was larger than 95%, when global analysis was undertaken. Similar dependences of reverse evolutionary rates on initial differentiation are observed in the hybrid (solid lines) and uncrossed populations (dotted lines), in both global and separate linear regressions.

rate of response throughout the experiment, or a two-phase linear model, with a first phase of response to selection followed by a plateau. The generation at which the second phase started was empirically determined²⁸. Adjusted R^2 , PRESS, Ellner and Turchin V_2 , and Mallows C_p were the statistics used to choose the best regression model^{29,30}. Heterogeneity among evolutionary trajectories was analysed using ANCOVA. These models used each separate character as the dependent variable, with generation defined as the covariate and evolutionary histories as a design factor with four categories: IO, ISO, ICO and IACO. Comparison of the rates of response under reverse evolution between uncrossed and hybrid populations were done by ANCOVA. In this model, the dependent variable was the slope estimated from one-phase linear regression models (using the response during reverse evolution) and the covariate was the estimated intercept of the regression model. Hybrid condition defined the factor in the ANCOVA, with uncrossed and hybrid categories. Power analyses were performed using the hybrid and uncrossed estimated least-squares means and standard deviations to estimate the power of t -tests for mean differences³⁰.

Received 25 May; accepted 18 September 2000.

- Muller, H. J. Reversibility in evolution considered from the standpoint of genetics. *Biol. Rev.* **14**, 261–280 (1939).
- Simpson, G. G. *The Major Features of Evolution* (Columbia Univ. Press, New York, 1953).
- Wright, S. *Evolution and the Genetics of Populations. Vol 3 Experimental Results and Evolutionary Deductions* (Univ. Chicago Press, Chicago, 1977).
- Bull, J. J. & Charnov, E. L. On irreversible evolution. *Evolution* **39**, 1149–1155 (1985).
- Gayon, J. *Darwinism's Struggle for Survival: Heredity and the Hypothesis of Natural Selection* (Cambridge Univ. Press, Cambridge and New York, 1998).
- Crill, W. D., Wichman, H. A. & Bull, J. J. Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* **154**, 27–37 (2000).
- Lenski, R. E. Experimental studies of pleiotropy and epistasis in *Escherichia coli*. II. Compensation for maladaptive effects associated with resistance to virus T4. *Evolution* **42**, 433–440 (1988).
- Rainey, P. B. & Travisano, M. Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69–72 (1998).
- Burch, C. L. & Chao, L. Evolution by small steps and rugged landscapes in the RNA virus $\phi 6$. *Genetics* **151**, 921–927 (1999).
- Moore, F. B.-G., Rozen, D. E. & Lenski, R. E. Pervasive compensatory adaptation in *Escherichia coli*. *Proc. R. Soc. Lond. B* **267**, 515–522 (2000).
- Levin, B. L., Perrot, V. & Walker, N. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in *Bacteria*. *Genetics* **154**, 985–997 (2000).
- Clarke, C. A., Mani, G. S. & Wynne, G. Evolution in reverse: clean air and the peppered moth. *Biol. J. Linn. Soc.* **26**, 189–199 (1985).
- Service, P. M., Hutchison, E. W. & Rose, M. R. Multiple genetic mechanisms for the evolution of senescence in *Drosophila melanogaster*. *Evolution* **42**, 708–716 (1988).
- Cook, L. M., Dennis, R. L. & Mani, G. S. Melanic morph frequency in the peppered moth in the Manchester area. *Proc. R. Soc. Lond. B* **266**, 293–297 (1999).
- Rose, M. R. Laboratory evolution of postponed senescence in *Drosophila melanogaster*. *Evolution* **38**, 1004–1010 (1984).
- Graves, J. L., Toolson, E. C., Jeong, C., Vu, L. N. & Rose, M. R. Desiccation, flight, glycogen, and postponed senescence in *Drosophila melanogaster*. *Physiol. Zool.* **65**, 268–286 (1992).
- Rose, M. R., Vu, L. N., Park, S. U. & Graves, J. L. Selection on stress resistance increased longevity in *Drosophila melanogaster*. *Exp. Gerontol.* **27**, 241–250 (1992).
- Fleming, J. E., Spicer, G. S., Garrison, R. C. & Rose, M. R. Two-dimensional protein electrophoretic analysis of postponed aging in *Drosophila*. *Genetica* **91**, 183–198 (1993).
- Tyler, R. H. et al. The effect of superoxide dismutase alleles on aging in *Drosophila*. *Genetica* **91**, 143–149 (1993).
- Leroi, A. M., Chippindale, A. K. & Rose, M. R. Long-term laboratory evolution of a genetic life-history trade-off in *Drosophila melanogaster*. I. The role of genotype-by-environment interaction. *Evolution* **48**, 1244–1257 (1994).
- Chippindale, A. K., Alipaz, J. A., Chen, H.-W. & Rose, M. R. Experimental evolution of accelerated development in *Drosophila*. I. Development speed and larval survival. *Evolution* **51**, 1536–1551 (1997).
- Fisher, R. A. *The Genetical Theory of Natural Selection* (Dover, New York, 1958).
- Reznick, D. N., Bryga, H. & Endler, J. A. Experimentally induced life-history evolution in a natural population. *Nature* **346**, 357–359 (1990).
- Coyne, J. A., Barton, N. H. & Turelli, M. Perspective: a critique of Sewall Wright's shifting balance theory of evolution. *Evolution* **51**, 643–671 (1997).
- Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K. & Rodriguez-Schettino, L. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**, 2115–2117 (1998).
- Cohan, F. M. & Hoffmann, A. A. Uniform selection as a diversifying force in evolution: evidence from *Drosophila*. *Am. Nat.* **134**, 613–637 (1989).
- Travisano, M., Mongold, J. A., Bennet, A. F. & Lenski, R. E. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267**, 87–90 (1995).
- SPSS, I. *SigmaPlot 5.0. Programming Guide* (SPSS Inc., Chicago, 1998).
- Glantz, S. A. & Slinker, B. K. *Primer of Applied Regression and Analysis of Variance* (McGraw-Hill Health Professions Division, New York, 1990).
- Zar, J. H. *Biostatistical Analysis* (Prentice Hall, Upper Saddle River, 1996).

Acknowledgements

We thank P. Beldade, A. K. Chippindale, L. D. Mueller, E. Sucena and particularly M. Matos for advice throughout the project and for comments on the manuscript. H.T. was supported by the Gulbenkian Foundation and Program PRAXIS XXI/FCT under the "Programa Gulbenkian de Doutorado em Biologia e Medicina".

Correspondence and requests for materials should be addressed to H.T. (e-mail: ht227@cam.ac.uk).

Integration of target and body-part information in the premotor cortex when planning action

Eiji Hoshi* & Jun Tanji*†

* Department of Physiology, Tohoku University School of Medicine, 2-1 Seiryō-cho, Aoba-ku, Sendai, 980-8575, Japan

† The Core Research for Evolutional Science and Technology Program, Kawaguchi, 332-0012, Japan

To plan an action, we must first select an object to act on and the body part (or parts) to use to accomplish our intention. To plan the motor task of reaching, we specify both the target to reach for and the arm to use. In the process of planning and preparing a motor task, information about the motor target and the arm to use must be integrated before a motor program can be formulated to generate the appropriate limb movement. One of the structures in the brain that is probably involved in integrating these two sets of information is the premotor area in the cerebral cortex of

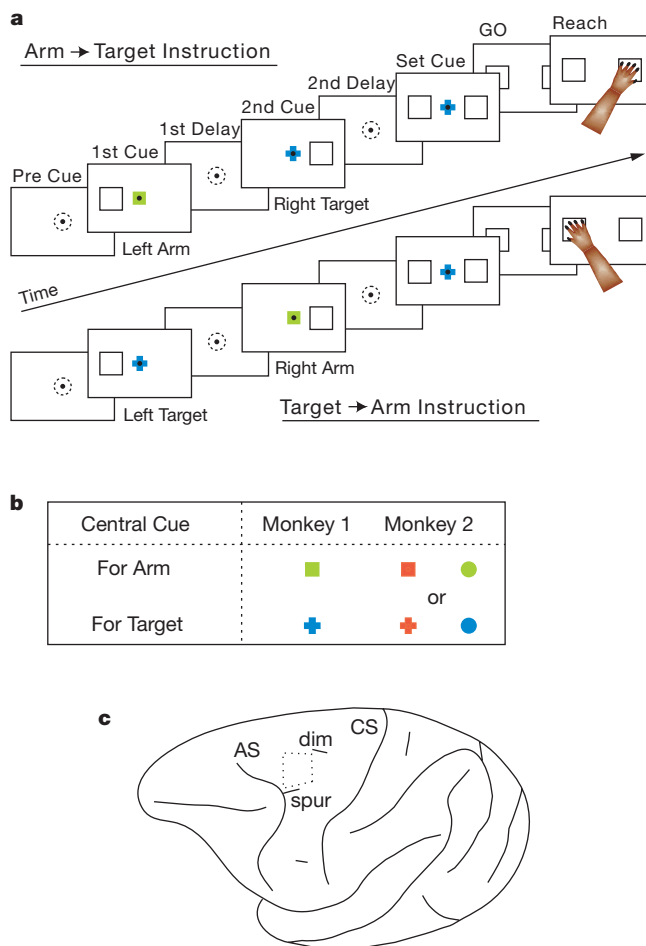


Figure 1 Experimental design. **a**, Temporal sequence of the behavioural events. Top row, a trial in which the two instructions were given in the order 'arm' then 'target'. Bottom row, a trial in which the two instructions were given in the order 'target' then 'arm'. The relative size of the fixation window is shown by the dotted circle. **b**, Central cues for monkeys 1 and 2. **c**, Cortical map of the recording sites. This work refers to the activity of neurons inside the dotted rectangle (PMD). AS, arcuate sulcus; CS, central sulcus; dim, superior precentral dimple; spur, spur of AS.

primates^{1–5}. The lateral sector of the dorsal premotor cortex is known to receive both visual and somatosensory input^{6–8}, and we show here that neurons in this area gather information about both the target and the body part, while subsequent activity specifies the planned action.

We trained two monkeys to perform a behavioural task that involved planning to reach for one of two targets with either the right or left arm (Fig. 1a; Methods). One set of cues specified which target (right or left) to reach for, and another specified which arm to use. The motor task started when the monkey gazed at a central

fixation point. The first and second instruction cues then appeared consecutively, each followed by a delay period. A central cue was superimposed on the fixation point to indicate whether the cue was arm- or target-related; if arm-related, a white square to the right or left of the fixation point instructed use of the right or left arm, respectively, while if the cue was target-related, the location of the white square relative to the fixation point corresponded to the right or left target. The colour and shape of this central cue, assigned as shown in Fig. 1b, determined whether the instruction cue was for the arm or the target. The order of the arm and target instructions was alternated in each block of 20 trials. For each instruction, right and left cues were presented pseudo-randomly. After the second delay period, a pair of squares (set cue) told the animals to get ready to reach for the target when the central fixation point disappeared, which was the 'go' signal. Thus, the monkeys were required to store the information given by the first cue, and then plan the target-reaching action after the second cue. Both monkeys performed the task with a success rate of 96–98%.

Here we focus on the neuronal activity during the first (first cue and delay period) and second (second cue and delay period) phases of the behavioural task. Although we recorded from the entire dorsal premotor cortex (PMd), here we report the activity in the postarcuate premotor cortex lateral to the superior precentral dimple and medial to the spur of the arcuate sulcus (Fig. 1c). This part of the PMd represents the forelimb^{8,9}, and it is where visual and somatosensory input converge^{6,7}. We found 214 task-related neurons that were active in the first or second phase (Mann–Whitney *U*-test).

During the first phase, we observed two types of neuronal activity of particular interest. One was selective activity that reflected which arm was to be used, as shown in Fig. 2a. In this example, activity was most pronounced when reaching was to be executed with the right arm (in RA1). The other was selective activity that reflected the

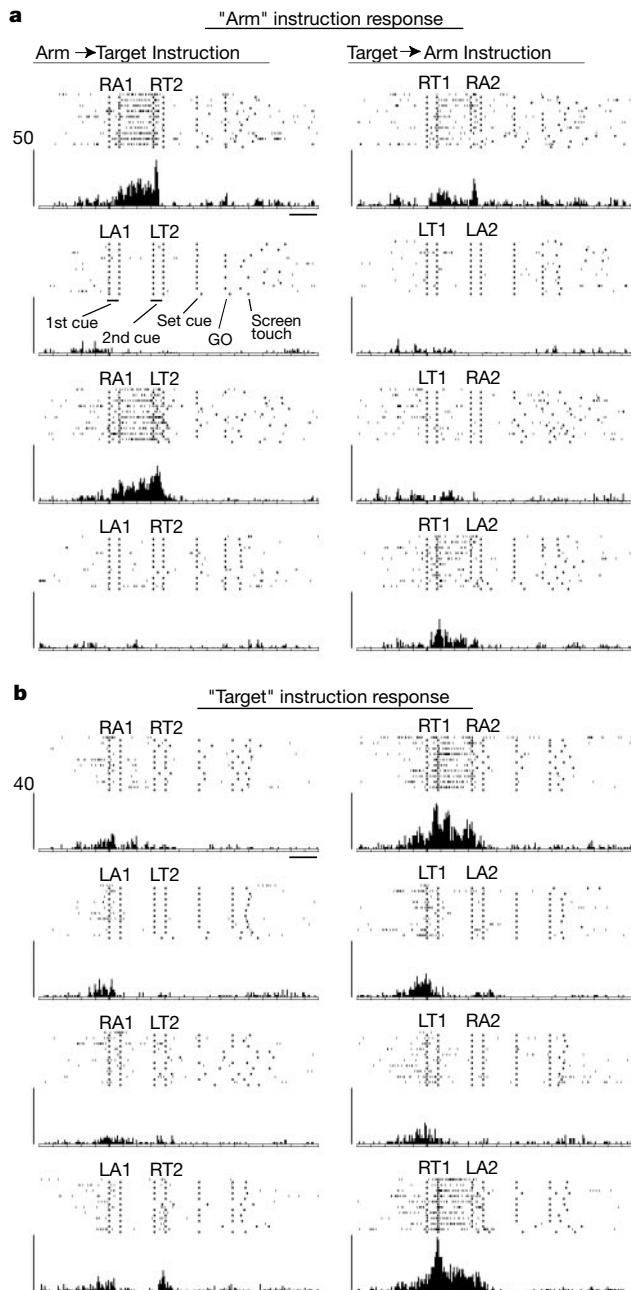


Figure 2 Two examples of neuronal activity in the PMd. **a**, Arm-selective activity. In rasters, each row represents a trial and each dot shows when the cell discharged. The seven marks under each trial represent when the task events appeared, as denoted in the left panel. The first and second instructions in each panel are shown on the top (RA, right arm; LA, left arm; RT, right target; LT, left target; 1, first cue; 2, second cue). In the perievent histograms (bin width, 40 ms), the ordinates represent the number of discharges per second. **b**, Target-selective activity. The rasters are aligned with the appearance of the first cue. Scale bar, 1 s.

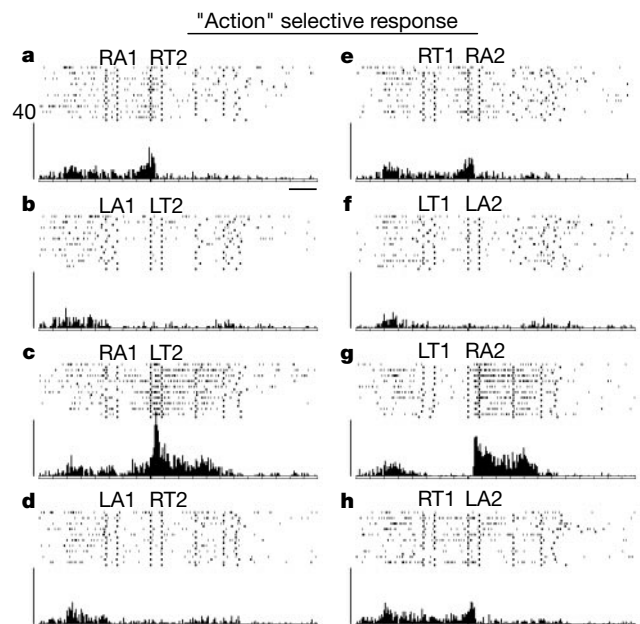


Figure 3 Action-selective activity. This neuron responded to the appearance of the second cue if the two instructions specified 'right arm' and 'left target' (**c** and **g**). The response reflected the combination of two cues rather than the second cue itself (GLM analysis; combination, $P < 1.0 \times 10^{-10}$; second cue, $P = 0.02$). Further analysis by two-way ANOVA revealed selectivity for both arm use ($P < 10^{-10}$) and target location ($P < 10^{-10}$). The activity was unaffected by the order of two instructions ($P > 0.01$, two-sample *t*-test). This neuron did not respond to first cues, or the second cues if the two cues were ordered as in **a**, **b**, **d**–**f** or **h**. The rasters are aligned with the appearance of the second cue.

location of the target, as shown in Fig. 2b (in RT1). Thus, these two types of neuronal activity reflect the arm to use or target to reach, rather than the physical position of the instruction cue (right or left of the white square). For 164 neurons active in the first phase, we investigated how many exhibited arm or target selectivity, and how many responded to the square position. We applied a two-way analysis of variance (ANOVA) (two factors: arm/target instruction of the first cue, left/right position of the white square), and classified them into five categories: (1) selective for arm- or target-instruction alone ($n = 23$); (2) selective for instruction and square-position ($n = 46$); (3) selective for square-position alone ($n = 64$); (4) non-selective ($n = 31$); and (5) non-responsive (that is, not task-related, $n = 50$). Of the 69 neurons that showed selectivity for arm/target instruction (categories (1) and (2)), 37 were selective for arm instruction, and 32 were selective for target instruction. Thus, information about arm use and target location was contained in these neurons during the neuronal activity in the first phase.

Second-phase activity was different from first-phase activity. A selective response to the second cue itself was observed in only a small number of neurons. Instead, most activity reflected a combination of the information given in the first and second cues. In the example shown in Fig. 3, the neuron was markedly active when the first cue instructed the right arm and the second cue instructed the left target (Fig. 3c). The same neuron was similarly active when the order of instructions was reversed (Fig. 3g). Thus, this neuron represents a specific action: reach for the left target with the right

arm. During the second phase, 156 neurons showed task-related activity. To investigate whether this activity reflected a combination of the information from the first and second cues or the second cue alone, we applied GLM (general linear model) analysis (see Methods). We found that 115 neurons reflected combined information, whereas only 4 were found to be selective for the second cue alone (Table 1). We also found that second-phase activity was unaffected by the order of the instructions (for arm or target, two-sample t -test) in 87% of the 115 neurons.

We then used a two-way ANOVA (two factors: left/right arm, left/right target) to examine whether activation of these 115 neurons reflected arm-use or target information alone, or was influenced by both; 63 (55%) were significant for both factors or for interaction between the two. These were called “action selective” neurons in this study (Table 1).

We further found that the information represented in the first phase was transformed in the second phase, sometimes even in a single neuron. The PMd neuron in the example shown in Fig. 4 was markedly active during the first phase after receiving left target information (LT1 in Fig. 4f and g). However, if the first instruction was for the right arm, the neuron did not respond at all when left target information was given in the second instruction (as shown during LT2 in Fig. 4c). During the second phase, the same neuron was particularly active when the instructions combined the left arm with the left target (Fig. 4b and f). Given this example, we studied how the 69 PMd neurons showing arm or target selectivity during

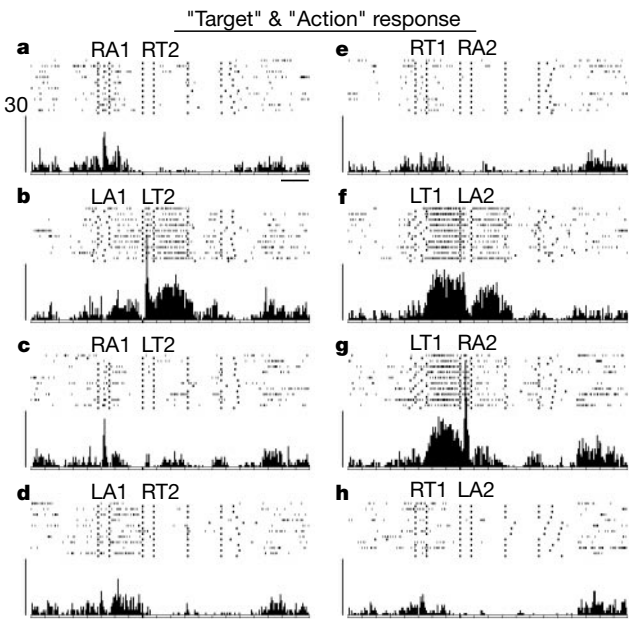


Figure 4 An example of neuronal activity exhibiting selectivity for a first cue instruction and also for a planned action. This neuron responded to the left target cue (LT1) during the first phase (**f** and **g**), but not necessarily to the LT2. During the second phase, the activity

reflected left-target reach with the left arm (**b** and **f**). This neuron was not active during the second phase when the first and second cues were ordered as in **a**, **c–e** or **h**.

Table 1 Relation of neuronal activity types between the first and second phases

Second phase	First phase					Total
	Instruction*	Instruction and position	Position†	Non-selective	Non-responsive	
Combination	10 (5)‡	17 (10)	25 (14)	10 (6)	35 (18)	97 (53)
Combination and second cue	0 (0)	4 (3)	5 (3)	5 (2)	4 (2)	18 (10)
Second cue	1	0	2	0	1	4
Non-selective	3	4	11	9	10	37
Non-responsive	9	21	21	7	–	58
Total	23 (5)	46 (13)	64 (17)	31 (8)	50 (20)	214 (63)

* Neurons selective for the arm/target instruction in the first phase.

† Neurons selective for the position of the white square in the first phase.

‡ The number of neurons further classified as ‘action selective’ in the second phase is shown in brackets.

the first phase behaved during the second phase. Thirty-nine were active during the second phase, and by applying GLM analysis we found that 31 of these exhibited selectivity for combined first- and second-cue information (Table 1). Of these, 18 were action selective, based on a two-way ANOVA looking for relatedness to the arm or target selected.

Could the action-selective neurons classified above have represented a combination of sensory cues (for example, a combination of left cue with green square and right cue with blue cross), rather than representing the forthcoming action? Two observations suggest that this interpretation is at least unlikely. First, we used two sets of central cues in monkey 2, as shown in Fig. 1b. Applying a two-way ANOVA (two factors: 2 sets of central cues, 4 movements), we found that none of the action-selective neuronal activity was significantly affected by the different central cues. Second, we added a control experiment with monkey 1, in which we extinguished the central cue (green square and blue cross) and presented only the white squares for the first and second instructions during a part of trials in a block. Because we removed the central cue in later trials mid-block, the monkey was able to perform the task correctly. In six of the neurons tested, none of the action-selective activity was altered by the absence of central cues.

We gave the monkeys two sets of instructions in two steps, thereby requiring them to plan and prepare a forthcoming action in two steps. We found 214 PMd neurons that were active in the first or second phase, of which 69 (32%) were active when arm-use or target-location information was given in the first instruction. When the second instruction was given, the activity in 115 neurons (54%) reflected combined information from the two instructions. In 63 (29%) of the 214 neurons, activity was selective for both the arm to use and the target to reach for (that is, for a future 'action'), indicating that the two sets of information converged in these neurons. The presence of neurons that represented both component information during the first phase and integrated information during the second phase seems to support the view that motor-target information and body-part information are integrated, at least in part, in the PMd.

The PMd has been implicated in motor programming^{10–17}, particularly in programming the parameters (for example, direction and amplitude) of motor output^{18,19}, and regional specificity within the PMd has been proposed^{5–8,20}. Neurons in a specialized portion of the PMd respond to both visual and somatosensory input, and neurons in the lateral sector of the PMd immediately medial to the arcuate spur have been found to respond to the presentation of visual objects placed either close to, or distant from, a subject monkey's body⁸. Anatomically, this particular portion of the PMd is known to receive visual input from the medial intraparietal area in the medial bank of the intraparietal sulcus, and somatosensory input from the parietal area 5 (refs 6, 7 and 21). Our findings extend knowledge of this specialized part of the PMd by showing how the neuronal activity reflects the process of integrating information about the target and the body-part to plan a forthcoming action. □

Methods

Animals, apparatus, and behavioural procedures

We used two monkeys (*Macaca fuscata*), which were cared for according to the NIH guidelines for the care and use of laboratory animals. We installed a colour monitor equipped with a touch-sensitive screen, and used methods described previously²² to record the activity of single neurons, eye position and muscle activity. We sampled the activity of arm, shoulder and back muscles bilaterally. The muscles showed movement-related activity, but no systematic changes in activity were detected in the other task periods. We identified cortical sulcal patterns, and measured the three-dimensional structure around the recording sites using an ultrasound imaging technique²³ (LOGIQ α system, GE Medical Systems). The monkeys were trained to perform a target-reaching task by following two sets of instructions while sitting in a chair. They were required to select the right or left target, and which arm to use to perform the task. The motor task started when the animal placed both hands in the start position, after an inter-trial interval of 3 s, and gazed at the fixation point (FP: 1.2° in diameter) that appeared in the centre of the monitor screen. If fixation was maintained for 1.2 s, the monkey was then given either the arm or

the target instruction: a white square (8° × 8°) that appeared to the left or right of the FP. A square to the left indicated the left arm (arm instruction) or left target (target instruction), whereas a square to the right indicated the right arm or right target. A small coloured cue (see Fig. 1b), which covered the central FP and appeared at the same time as the white square, showed whether the instruction was for the arm or the target. The first instruction (400 ms) was followed by the first delay period (≥1.2 s), then by the second instruction (400 ms) and the second delay (≥1.2 s), after which a square appeared on both sides of the FP (set cue), telling the animal to get ready to reach for the correct target in response to the disappearance of the FP (the 'go' signal); subsequent reaching for the target with a reaction time < 1 s was rewarded with fruit juice. During the delay periods, the animals had to maintain fixation on the FP to proceed to the next phase of the task. Fixation was broken after the appearance of the first or second cues in 21–26% of trials (see Supplementary Information for eye-position traces during task performance). We analysed the effects of eye movement during this period, and found that neuronal activity during the delay periods was unaffected when monkeys made saccades to the cue location (Mann–Whitney *U*-test). The order of appearance of the arm and target instructions was alternated in a block of 20 trials. A series of five 250-Hz tones after a reward signalled reversal of the order.

Data analysis

Our database included only neurons whose activity was recorded during more than two blocks of trials in each sequence of instruction cues. The significance level of all statistical tests was set at $P < 0.01$. We analysed neuronal activity during two phases of the task: from the appearance of the first cue to that of the second, and from the appearance of the second cue to that of the set cue. This work relates to 214 PMd neurons that showed task-related activity in the first or second phase. A neuron was judged task-related if its discharge rate differed significantly (Mann–Whitney *U*-test, compared in each of eight trial types, corrected for multiple comparisons) from that of the control period (1,000 ms before FP onset) during a given time window. The default time window was the last 1,000 ms of each phase. For 21 neurons, the time window was set from 100 ms to 1,100 ms after the cue onset. We performed a two-way ANOVA of the task-related activity in the first phase, looking at the position of the white square and the type of instruction (target or arm). We applied a GLM analysis (GLM for two categories: four second cues, four combinations of the first and second cues) to the activity in the second phase. We used the equation: activity = $\beta_0 + \beta_1(\text{second cue}) + \beta_2(\text{combination})$, where β_1 and β_2 are the slopes of the best-fit lines and β_0 is the intercept.

Received 20 June; accepted 11 September 2000.

- Dum, R. P. & Strick, P. L. In *Motor Control: Concepts and Issues* (eds Humphrey, D. R. & Freund, H.-J.) 383–397 (Wiley & Sons, New York, 1991).
- Wise, S. P. The primate premotor cortex: past, present, and preparatory. *Annu. Rev. Neurosci.* **8**, 1–19 (1985).
- Passingham, R. E. *The Frontal Lobes and Voluntary Action* (Oxford Univ. Press, Oxford, 1993).
- Kurata, K. Information processing for motor control in primate premotor cortex. *Behav. Brain Res.* **61**, 135–142 (1994).
- Rizzolatti, G., Luppino, G. & Matelli, M. The organization of the cortical motor system: new concepts. *Electroencephalogr. Clin. Neurophysiol.* **106**, 283–296 (1998).
- Caminiti, R., Ferraina, S. & Johnson, P. B. The sources of visual information to the primate frontal lobe: a novel role for the superior parietal lobule. *Cereb. Cortex* **6**, 319–328 (1996).
- Matelli, M., Govoni, P., Galletti, C., Kutz, D. F. & Luppino, G. Superior area 6 efferents from the superior parietal lobule in the macaque monkey. *J. Comp. Neurol.* **402**, 327–352 (1998).
- Fogassi, L. et al. Visual responses in the dorsal premotor area F2 of the macaque monkey. *Exp. Brain Res.* **128**, 194–199 (1999).
- Kurata, K., Okano, K. & Tanji, J. Distribution of neurons related to a hindlimb as opposed to forelimb movement in the primate premotor cortex. *Exp. Brain Res.* **60**, 188–191 (1985).
- Halsband, U. & Passingham, R. E. The role of premotor and parietal cortex in the direction of action. *Brain Res.* **240**, 368–372 (1982).
- Passingham, R. E. Premotor cortex and preparation for movement. *Exp. Brain Res.* **70**, 590–596 (1988).
- Petrides, M. Deficits in non-spatial conditional associative learning after periarculate lesions in the monkey. *Behav. Brain Res.* **16**, 95–101 (1985).
- Kurata, K. & Wise, S. P. Premotor cortex of rhesus monkeys: set-related activity during two conditional motor tasks. *Exp. Brain Res.* **69**, 327–343 (1988).
- Mushiake, H., Inase, M. & Tanji, J. Neuronal activity in the primate premotor, supplementary, and precentral motor cortex during visually guided and internally determined sequential movements. *J. Neurophysiol.* **66**, 705–718 (1991).
- Boussaoud, D. & Wise, S. P. Primate frontal cortex: neuronal activity following attentional versus intentional cues. *Exp. Brain Res.* **95**, 15–27 (1993).
- di Pellegrino, G. & Wise, S. P. Visuospatial versus visuomotor activity in the premotor and prefrontal cortex of a primate. *J. Neurosci.* **13**, 1227–1243 (1993).
- Shen, L. & Alexander, G. E. Preferential representation of instructed target location versus limb trajectory in dorsal premotor area. *J. Neurophysiol.* **77**, 1195–1212 (1997).
- Riehle, A. & Requin, J. Monkey primary motor and premotor cortex: single-cell activity related to prior information about direction and extent of an intended movement. *J. Neurophysiol.* **61**, 534–549 (1989).
- Kurata, K. Premotor cortex of monkeys: set- and movement-related activity reflecting amplitude and direction of wrist movements. *J. Neurophysiol.* **69**, 187–200 (1993).
- Wise, S. P., Boussaoud, D., Johnson, P. B. & Caminiti, R. Premotor and parietal cortex: corticocortical connectivity and combinatorial computations. *Annu. Rev. Neurosci.* **20**, 25–42 (1997).
- Johnson, P. B., Ferraina, S., Bianchi, L. & Caminiti, R. Cortical networks for visual reaching: physiological and anatomical organization of frontal and parietal lobe arm regions. *Cereb. Cortex* **6**, 102–119 (1996).
- Hoshi, E., Shima, K. & Tanji, J. Neuronal activity in the primate prefrontal cortex in the process of motor selection based on two behavioral rules. *J. Neurophysiol.* **83**, 2355–2373 (2000).

23. Tokuno, H., Hatanaka, N., Takada, M. & Nambu, A. B-mode and color Doppler ultrasound imaging for localization of microelectrode in monkey brain. *Neurosci. Res.* 36, 335–338 (2000).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Kurama, Y. Takahashi and S. Hoffman for technical assistance. This work was supported by the Japan Society for the Promotion of Science (E.H.), by the Ministry of Education, Science, and Culture of Japan, and by the Japan Science and Technology Corporation (J.T.).

Correspondence and requests for materials should be addressed to J.T. (e-mail: tanji@mail.cc.tohoku.ac.jp).

MOD-1 is a serotonin-gated chloride channel that modulates locomotory behaviour in *C. elegans*

Rajesh Ranganathan*, Stephen C. Cannon† & H. Robert Horvitz*

* Howard Hughes Medical Institute, Department of Biology, Room 68-425, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

† Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, and Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

The neurotransmitter and neuromodulator serotonin (5-HT) functions by binding either to metabotropic G-protein-coupled receptors (for example, 5-HT₁, 5-HT₂, 5-HT₄ to 5-HT₇), which mediate 'slow' modulatory responses through numerous second messenger pathways¹, or to the ionotropic 5-HT₃ receptor, a non-selective cation channel that mediates 'fast' membrane depolarizations². Here we report that the gene *mod-1* (for modulation of locomotion defective) from the nematode *Caenorhabditis elegans* encodes a new type of ionotropic 5-HT receptor, a 5-HT-gated chloride channel. The predicted MOD-1 protein is similar to members of the nicotinic acetylcholine receptor family of ligand-gated ion channels, in particular to GABA (γ -aminobutyric acid)- and glycine-gated chloride channels. The MOD-1 channel has distinctive ion selectivity and pharmacological properties. The reversal potential of the MOD-1 channel is dependent on the concentration of chloride ions but not of cations. The MOD-1 channel is not blocked by calcium ions or 5-HT_{3A}-specific antagonists but is inhibited by the metabotropic 5-HT receptor antagonists mianserin and methiothepin. *mod-1* mutant animals are defective in a 5-HT-mediated experience-dependent behaviour³ and are resistant to exogenous 5-HT, confirming that MOD-1 functions as a 5-HT receptor *in vivo*.

The locomotory rate of a well-fed *C. elegans* hermaphrodite slows when it encounters bacteria (the basal slowing response³); this slowing response is enhanced markedly if the animal has been deprived of food for 30 min before the encounter (the enhanced slowing response³ (Fig. 1a). In a genetic screen for mutants defective in the enhanced slowing response, we identified the gene *mod-1*, which is defined by the mutation *n3034* (ref. 3). On Petri dishes with bacteria, the locomotory rate of food-deprived *mod-1* mutants is significantly faster than that of the wild type (strain N2) (Fig. 1a, grey bars). On dishes without bacteria, the locomotory rate of food-deprived *mod-1* mutants is not different from that of food-deprived wild-type animals (Fig. 1a, grey bars). Well-fed *mod-1* mutants

show no defect in their basal slowing response to bacteria (Fig. 1a, black bars).

The swimming behaviour of *mod-1(n3034)* mutants was not affected by exogenous 5-HT treatment, whereas exogenous 5-HT inhibits wild-type *C. elegans* locomotion⁴; and this phenotype was semi-dominant in a time-dependent manner (Fig. 1b), which aided our cloning of the *mod-1* gene (see below). This 5-HT resistance of *mod-1* mutants is consistent with the finding that the enhanced slowing response is mediated by 5-HT signalling³ and the hypothesis that *mod-1* mutants are deficient in 5-HT signalling³.

We used standard genetic three-factor mapping and scored the 5-HT resistance of *mod-1* mutants to locate *mod-1* to a ~400 kilobase (kb) region on chromosome V. Cosmid C38E12 from this region rescued the recessive aspect of the *Mod-1* mutant phenotype of 5-HT resistance, as assayed by the response to exogenous 5-HT after 5 min (Fig. 1c). We narrowed the rescuing activity to a 5.5-kb fragment (Fig. 1c) containing a single predicted gene⁵. We isolated several complementary DNAs with SL1 trans-spliced leaders, located at the 5' ends of many *C. elegans* transcripts⁶, and determined the complete *mod-1* transcription unit (Fig. 2a). Genomic subclones lacking all or part of this transcription unit did not rescue the *Mod-1* phenotype of 5-HT resistance (Fig. 1c).

MOD-1 is similar to members of the nicotinic acetylcholine receptor (nAChR) family of ligand-gated ion channels, in particular to GABA- and glycine-gated chloride channels (Fig. 2b). We found a single-base transition mutation in the *mod-1* coding sequence in *n3034* mutants (Fig. 2b, c), in which Ala 281 (codon GCT) is changed to valine (GTT) in the predicted M2 transmembrane domain of the MOD-1 protein. M2 is thought to be critical for channel function. We used site-directed mutagenesis to introduce this C-to-T (A281V) mutation into the 5.5-kb minimal rescuing

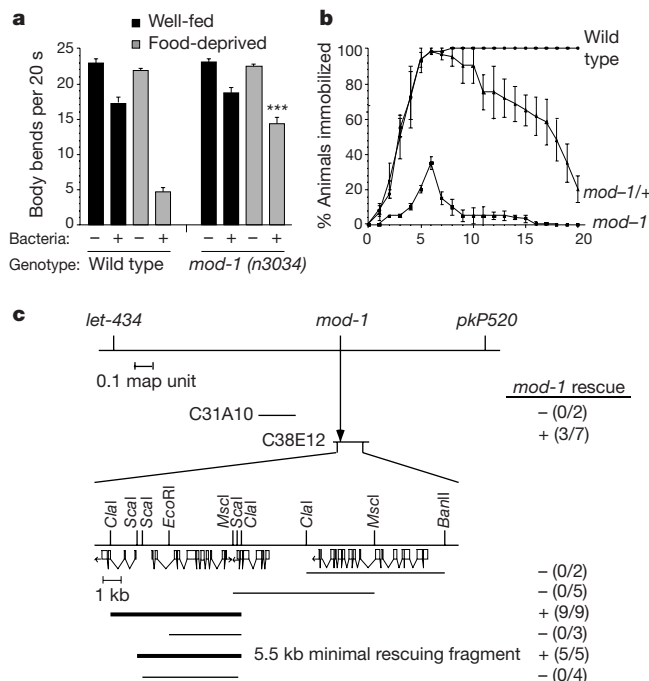


Figure 1 Phenotypic characterization and cloning of *mod-1*. **a**, *mod-1(n3034)* mutants were defective in the enhanced slowing locomotory response exhibited by food-deprived wild-type animals³. Ten trials were performed for each genotype. Error bars represent s.e.m.; asterisks, $P < 0.0001$, Student's *t*-test. **b**, Time course of response to 33 mM 5-HT. The resistance to exogenous 5-HT of *mod-1(n3034)* mutants was recessive at early time points and was semi-dominant at later time points. **c**, Genetic and physical maps of the relevant region of linkage group V. Cosmid C38E12 and subclones shown in bold rescued *mod-1*; +, rescue; -, no rescue. Numbers in parentheses represent the fraction of transgenic lines tested that rescued *mod-1*.

fragment and generated transgenic animals carrying extrachromosomal arrays of this fragment, Ex[MOD-1(A281V)]. These transgenic animals displayed resistance to exogenous 5-HT (Fig. 3a), confirming that the A281V mutation in MOD-1 is sufficient to cause 5-HT resistance.

To determine the effect of eliminating *mod-1* function, we analysed two deletion alleles of *mod-1*, *nr2043* (ref. 7) and *ok103* (Fig. 2a), obtained by screening libraries of mutagenized animals for large deletions in the *mod-1* genomic locus⁸. Both *mod-1(nr2043)* and *mod-1(ok103)* mutants, when deprived of food, were defective in the enhanced slowing response (Fig. 3b), whereas neither was defective in the basal slowing response (data not shown). Both deletion mutants were resistant to exogenous 5-HT (Fig. 3a). The 5-HT resistance caused by the two deletion alleles was completely recessive throughout the 20-min time course (Fig. 3a; and data not shown), consistent with our observation that animals heterozygous for large chromosomal deficiencies that uncover the *mod-1* genomic locus are not 5-HT-resistant (data not shown). The molecular nature of the *mod-1(nr2043)* and *mod-1(ok103)* mutations suggests that they are null alleles. As null alleles confer the same phenotype as Ex[MOD-1(A281V)], *mod-1(nr2043)* is likely to be a dominant-negative allele.

MOD-1 protein is predicted to contain a large extracellular amino terminus with two cysteine residues separated by 13 amino acids, a region that is conserved in the nAChR family and that may have a role in subunit assembly and insertion into the membrane⁹ (Fig. 2b). There are four predicted transmembrane regions (M1–M4) and a large cytoplasmic domain between M3 and M4 (Fig. 2b). The M2 domain may participate in forming the pore of the channel after these nAChR family members assemble as multimeric complexes in

the cell membrane¹⁰. In other family members, the residues in and around the M2 domain help establish ion selectivity¹⁰. We could not confidently predict the ion selectivity of the MOD-1 channel from sequence comparisons.

We examined the ability of MOD-1 to form functional channels in voltage-clamped *Xenopus laevis* oocytes. Oocytes injected with *mod-1* complementary RNA and voltage-clamped at –70 mV responded to 1 μ M 5-HT with a large, rapidly developing, non-oscillatory inward current (Fig. 4a), whereas control oocytes injected with water showed no such response to application of any agonist (data not shown). Even 1,000 μ M glycine, acetylcholine, GABA, glutamate or histamine—neurotransmitters that can activate ion channels^{11–13}—did not elicit an inward current in *mod-1*-injected oocytes (Fig. 4a; and data not shown). Similarly, 10 μ M ivermectin failed to elicit any current in MOD-1-injected oocytes (data not shown). Ivermectin, an anthelmintic drug that activates glutamate-gated chloride channels in nematodes¹⁴, can activate a rat $\alpha 1\beta 2\gamma 2$ S-GABA-gated chloride channel¹⁵ and enhances acetylcholine-evoked responses of the human nAChR $\alpha 7$ receptor¹⁶. Dopamine, octopamine or tyramine did not elicit MOD-1-dependent inward currents at 100 μ M, but elicited small responses at millimolar concentrations (data not shown). *mod-1* mutants did not have defects associated with impaired dopamine or octopamine signalling (ref. 3; and M. Alkema, R.R. and H.R.H., unpublished observations). We conclude that 5-HT is most likely to be the native ligand for the MOD-1 channel. 5-HT dose–response experiments using oocytes showed that the MOD-1 channel has an effector concentration for half-maximal response (EC_{50}) of $1.0 \pm 0.1 \mu$ M (Fig. 4b), which is lower than that of the human 5-HT_{3a} channel

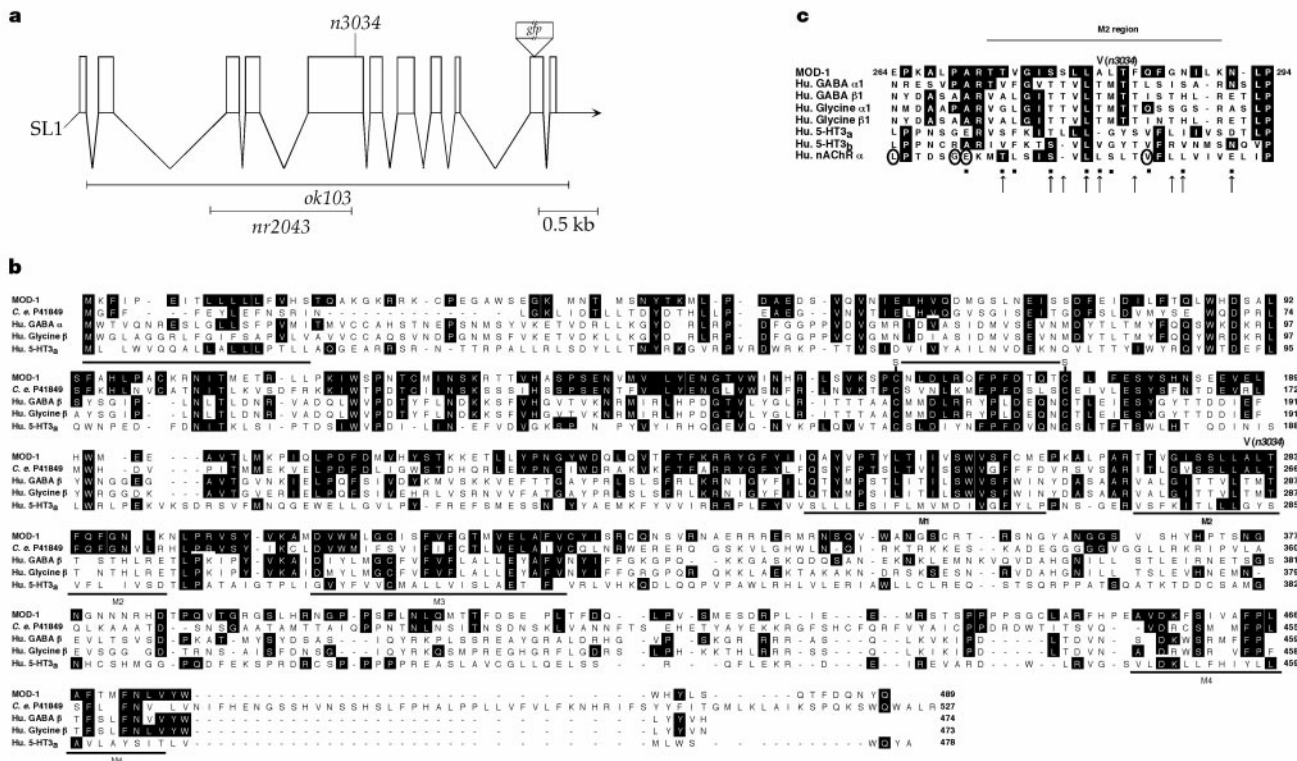


Figure 2 Sequence analysis of *mod-1*. **a**, *mod-1* intron–exon structure. Open boxes, coding regions; lines, untranslated regions; arrow, direction of transcription; *SL1*, *SL1* trans-spliced leader. The *ok103* and *nr2043* deletions are depicted (see Supplementary Information), and the positions of the *n3034* mutation and insertion site of the *gfp* gene are indicated. **b**, Amino-acid sequence alignment of MOD-1 with a *C. elegans* predicted protein (P41849) that is most similar to MOD-1, human GABA_A receptor $\beta 1$ subunit (A40336), human glycine receptor β subunit (NP_000815.1), and the human 5-HT_{3a} subunit (BAA08387). The putative signal sequence and M1–M4 regions are underlined,

and the Cys–Cys loop is indicated (S–S). Amino acids conserved between MOD-1 and at least one other protein are shown in black boxes. *mod-1(n3034)* is a C-to-T mutation resulting in an A281V substitution. **c**, Alignment of the M2 regions from MOD-1, human GABA_A $\alpha 1$ subunit (A60652), human GABA_A $\beta 1$ subunit, human glycine $\alpha 1$ subunit (NP_000162), human glycine $\beta 1$ subunit, 5-HT_{3a}, 5-HT_{3b} (AF080582) and human nAChR α -subunit (P02708). Dots, residues that probably face the pore in the nAChR¹⁰; arrows, residues that probably face the pore in the GABA_A receptor¹⁰; circles, residues conserved in cation channels that are not conserved in MOD-1.

($2.9 \pm 0.1 \mu\text{M}$)¹⁷. Unlike the 5-HT_{3a} channel², the MOD-1 channel was not blocked by calcium (Fig. 4c). High concentrations of granisetron (Fig. 4d) and ondansetron (data not shown), both of which are potent antagonists of the 5-HT_{3a} channel¹⁸, did not affect the action of 5-HT on MOD-1 channels.

Pretreatment of wild-type *C. elegans* with mianserin or methiothepin—5-HT receptor antagonists—prevents food-deprived animals from exhibiting the wild-type enhanced slowing response³. Therefore, even though both compounds have so far been considered to be primarily antagonists of metabotropic 5-HT

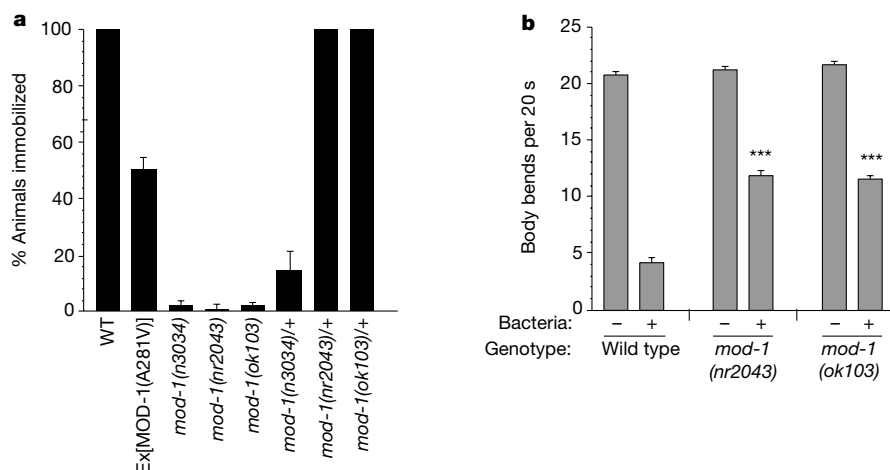


Figure 3 *mod-1* deletion alleles caused 5-HT resistance and defects in the enhanced slowing response. **a**, 5-HT sensitivity, scored after 20 min in 33 mM 5-HT. Ex[MOD-1(A281V)] is an extrachromosomal array containing the wild-type *lin-15* gene and the *mod-1* genomic locus encoding the A281V mutant form of the MOD-1 protein. Control

transgenic lines with extrachromosomal arrays containing wild-type *lin-15* and *mod-1* genes were not resistant to 5-HT (data not shown). **b**, *mod-1* deletion mutants were defective in the enhanced slowing response. Ten trials were performed for each genotype. Error bars represent s.e.m.; asterisks, $P < 0.0001$, Student's *t*-test.

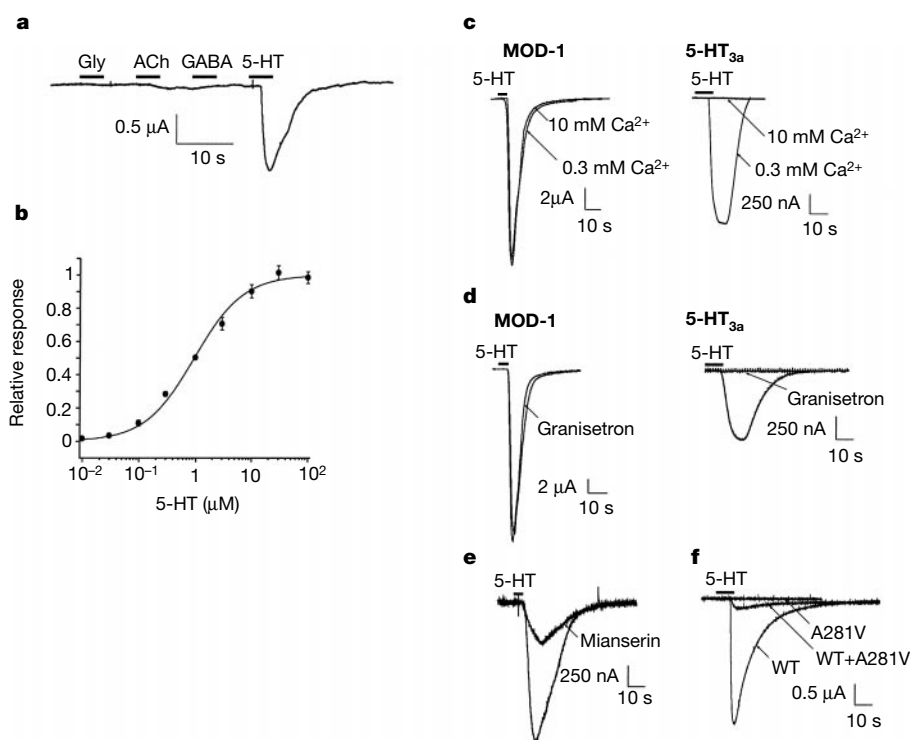


Figure 4 MOD-1 is a 5-HT-gated ion channel distinct from the 5-HT_{3a} channel. Recordings were from *Xenopus* oocytes voltage clamped at -70 mV in standard ND-96 bath solution, except in **c** where the Ca^{2+} concentrations were varied. Bars, durations of agonist application. **a**, 1 μM 5-HT elicited a rapidly developing inward current from oocytes injected with *mod-1* cRNA. A representative trace is shown from an oocyte treated with 1 mM Gly, ACh and GABA, and 1 μM 5-HT in ND-96 ($n = 5$). The order of agonist application was not important (data not shown). **b**, 5-HT dose response curve; $\text{EC}_{50} = 1.0 \pm 0.1 \mu\text{M}$; $n \geq 4$ for each concentration; error bars represent s.e.m. **c**, Extracellular calcium did not block the MOD-1 channel. Representative traces are shown from oocytes injected with *mod-1* ($n = 3$) or 5-HT_{3a} ($n = 2$) cRNA and treated with 1 μM (*mod-1*) or 10 μM (5-HT_{3a}) 5-HT. **d**, The 5-HT_{3a}-specific blocker granisetron did not block MOD-1.

Representative traces are shown from oocytes injected with *mod-1* or 5-HT_{3a} cRNA ($n = 2$ for each) and treated with 1 μM (*mod-1*) or 10 μM (5-HT_{3a}) 5-HT with or without 130 μM granisetron. **e**, Mianserin blocked the MOD-1 channel. Representative traces are shown from *mod-1*-injected oocytes ($n = 5$) treated with 0.5 μM 5-HT with or without 50 μM mianserin. **f**, The MOD-1(A281V) mutant channel has a dominant-negative effect. Representative traces are shown from oocytes injected with wild-type *mod-1* cRNA (WT), *mod-1* cRNA encoding the A281V mutant protein (A281V), or four parts wild type to one part A281V (WT+A281V) ($n \geq 4$ for each class). Wild-type oocytes were treated with 1 μM 5-HT; A281V and WT+A281V oocytes were treated with 10 μM 5-HT; 1 μM 5-HT did not elicit any response from either A281V or WT+A281V oocytes.

receptors^{19,20}, we tested their effects on MOD-1 in oocytes. The MOD-1 channel was inhibited by mianserin and methiothepin, with inhibition constants (K_i) of $\sim 19 \mu\text{M}$ and $\sim 32 \mu\text{M}$, respectively (Fig. 4e; and data not shown). Pretreatment of *mod-1* mutants with mianserin or methiothepin did not further affect the defective enhanced slowing response of these animals (data not shown). Mianserin and methiothepin may therefore interfere with the enhanced slowing response of *C. elegans* by antagonizing the MOD-1 5-HT-gated channel.

We tested the effect on channel function of the MOD-1(A281V) substitution in the *mod-1(n3034)* mutant. When antisense RNA (cRNA) encoding MOD-1(A281V) was injected into oocytes, no 5-HT-gated responses were observed (Fig. 4f). When the mutant cRNA was co-injected with roughly a fourfold excess of wild-type *mod-1* cRNA, the magnitude of the current through the wild-type channels was markedly reduced compared with that of oocytes that had been injected with the same amount of only the wild-type cRNA (Fig. 4f). These findings indicate that the MOD-1 channel may be multimeric, and that mutant MOD-1(A281V) channel subunits may interfere in a dominant manner with the function of wild-type MOD-1 channel subunits. Alternatively, MOD-1(A281V) channel subunits might interfere with wild-type MOD-1 channel function in a nonspecific manner, for example, by affecting assembly, transport or stability of all membrane proteins. Our genetic data support this hypothesis in that *mod-1(n3034)* encodes a dominant-negative form of MOD-1.

Current–voltage (I – V) relationships showed that the 5-HT-dependent current from MOD-1-expressing oocytes reverses

direction (in ND-96 bath solution) near -20 mV (data not shown), significantly different from the near 0 mV reversal potential of the 5-HT_{3a} non-selective cation channel (ref. 2; and our unpublished data). The chloride reversal potential (E_{Cl}) in *Xenopus* oocytes is roughly -20 mV (ref. 21), consistent with the hypothesis that MOD-1 is a chloride channel. Moreover, when external sodium chloride was replaced with choline chloride, there was no change in the MOD-1 reversal potential (data not shown), indicating that sodium was not the primary charge carrier through the MOD-1 channel.

We directly tested whether chloride-ion concentration affected the reversal potential of the MOD-1 channel. To avoid activities of chloride channels endogenous to the *Xenopus* oocyte²¹ and to be able to manipulate the concentrations of ions on both sides of the membrane, we conducted these experiments using human embryonic kidney 293 (HEK293) cells transiently transfected with a *mod-1* cDNA. We found that the peak I – V relationship showed no evidence of rectification (Fig. 5a), indicating that there was no voltage-dependent block by divalent cations or small organic molecules. Mock-transfected control HEK293 cells did not have any 5-HT-gated responses (data not shown). We also found that a 10-fold change in external chloride concentration shifted the reversal potential by 53 mV (Fig. 5b), in good agreement with the theoretical shift of 58 mV predicted by the Nernst equation if the MOD-1 channel were perfectly selective for chloride ions. Thus, MOD-1 is a 5-HT-gated chloride channel.

As the M2 region of the MOD-1 channel is similar to those of GABA- and glycine-gated chloride channels (Fig. 2c) and these

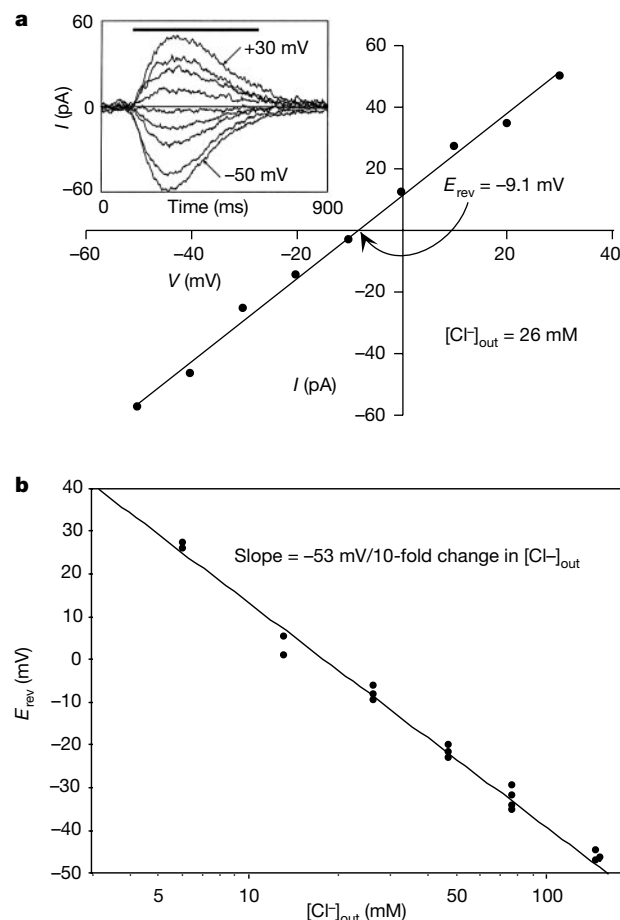


Figure 5 MOD-1 is a 5-HT-gated chloride channel. Whole-cell recordings from HEK293 cells transfected with a *mod-1* cDNA. **a**, The I – V relationship of the maximal responses from a representative cell when the external solution contained 120 mM sodium gluconate and 20 mM NaCl, resulting in $[\text{Cl}^-]_{\text{out}} = 26 \text{ mM}$, $[\text{Na}^+]_{\text{in}} = 0$, and $[\text{Cl}^-]_{\text{in}} = 12 \text{ mM}$

(see Methods for complete salt compositions). $E_{\text{rev}} = -9.1 \text{ mV}$. Inset, raw traces for this cell at each of the nine voltage steps from -50 mV to $+30 \text{ mV}$. Bar, duration of 100 mM 5-HT application. **b**, The reversal potential varied linearly with $\log[\text{Cl}^-]_{\text{out}}$. Each data point represents E_{rev} of a single cell at one particular $[\text{Cl}^-]_{\text{out}}$ as in **a**.

channels are permeant to anions other than chloride²², we determined whether MOD-1 channels were also permeant to several anions. We measured the MOD-1 reversal potential in *Xenopus* oocytes, substituting chloride ions in the external medium with an equivalent concentration of six other anions of various sizes. The rank order of anion permeability for MOD-1 channels was thiocyanate ($E_{rev} < -40$ mV) > bromide $\approx$ iodide > chloride ($E_{rev} \approx -20$ mV) $\gg$ fluoride $\gg$ aspartate $\gg$ gluconate ($E_{rev} > +50$ mV) (data not shown). This order is in agreement with the rank order for GABA- and glycine-gated chloride channels²², indicating that the permeation properties of the MOD-1 channel are similar to those of GABA- and glycine-gated chloride channels.

We generated a green fluorescent protein (GFP) reporter of *mod-1* by inserting the *gfp* gene²³ into the *mod-1* minimal genomic rescuing fragment between the M3 and M4 domains (Fig. 2a). Stable chromosomally integrated lines of this construct²⁴ were rescued for the *mod-1* mutant phenotype (data not shown). *mod-1::GFP* reporter expression was observed in the cell bodies and axons (the latter presumably as a consequence of reporter overexpression) of several neurons in the head, ventral cord and tail of the animal (data not shown). No reporter expression was observed in any muscle cells. These observations indicate that the MOD-1 5-HT-gated chloride channel probably functions in neurons in *C. elegans*.

Voltage-clamp recordings from isolated embryonic Retzius cells from the medicinal leech *Hirudo medicinalis* indicate that 5-HT applications can result in hyperpolarizing chloride currents²⁵. Recordings from dialysed whole cells and outside-out patches from these Retzius cells suggest that the 5-HT-gated chloride current is a consequence of a direct gating of a channel rather than the output of an indirect mechanism requiring second messenger cascades²⁶. Our molecular identification and characterization of MOD-1 support this conjecture and demonstrate directly the existence of a 5-HT-gated chloride channel. We show that this channel acts as a 5-HT receptor *in vivo* and that its function is necessary for a 5-HT-mediated experience-dependent modulation of behaviour. If mammalian counterparts of MOD-1 exist, their identification and characterization might provide insights concerning the myriad neurobiological effects of 5-HT and define new targets for the development of human pharmaceuticals. □

Methods

Mapping and cloning of *mod-1*

We mapped *mod-1* using standard procedures²⁷ (see Supplementary Information). We performed germline transformation experiments²⁸ by injecting the various constructs along with 80 μ g ml⁻¹ pL15EK (which contains the wild-type *lin-15* gene) into a *mod-1*(n3034); *lin-15*(n765ts) strain and scoring stable transgenic lines, which produced non-Lin progeny. Rescued lines had at least half of the animals exhibiting normal sensitivity to 5-HT at 5 min in assays of 5-HT resistance.

Behavioural assays

Locomotor activity was assayed as described³. In Figs 1a and 3b, each trial involved testing at least five animals for each of the conditions; a given animal was tested for only one condition. *P* values were calculated by comparing the combined data for the mutants from all the separate trials to the combined data for the wild-type animals assayed in parallel for each condition of each separate trial. To assay 5-HT resistance, we placed 20 animals in 200 μ l of 33 mM 5-HT (creatinine sulphate salt; Sigma) dissolved in M9 buffer²⁷ in 96-well microtitre wells and scored the swimming behaviour of such animals as active or immobile every minute for 20 min for the time-course experiments, at 5 min for germline transformation rescue experiments, or at 20 min (Fig. 3a). In this assay, an animal was scored as immobile if it did not exhibit any swimming motion for a period of 5 s.

Electrophysiological studies of MOD-1

An *EcoRI*/*EcoRI* fragment containing a full-length *mod-1* cDNA, including the 5' and 3' UTRs, was cloned into an *EcoRI* site of the vector pGEMHE²⁹ for oocyte expression and into the *EcoRI* site of vector GW1-CMV (British Biotech) for transfection into HEK293 (ATCC # CRL-1533) cells. Capped cRNA for the oocytes was synthesized as described²⁹ and dissolved in water. RNA concentrations were determined by agarose gel electrophoresis and absorption spectroscopy. *Xenopus* oocytes were collected and injected with 50 nl RNA as described²⁹, and two-electrode voltage-clamp recordings were performed 1–3 days later at room temperature (22–25 °C) as described²⁹. All compounds and wash solutions were applied to oocytes using a gravity-assisted perfusion system. The standard bath solution for the agonist, antagonist, and dose–response experiments was ND-96

(in mM): 96 NaCl, 2 KCl, 0.3 CaCl₂, 1 MgCl₂, 5 HEPES (pH 7.6). In Fig. 4b, each oocyte was sequentially subjected to 5-s treatments of increasing concentrations of 5-HT with 30 s of wash between applications. This protocol was continued from 10 nM to 1 μ M, and then only one concentration higher than 1 μ M was tested on any particular oocyte, as concentrations greater than 1 μ M 5-HT led to desensitization and a decrease in the maximal response elicited by subsequent 5-HT applications. Responses from separate oocytes were combined by normalizing the amplitudes of each oocyte's responses to the various 5-HT concentrations to that elicited by 1 μ M 5-HT. The normalized data were fit to the function $y = V_{max}[x^n/(k^n + x^n)]$ with a Hill coefficient (n) = 1, and then re-normalized to a maximal value of 1 by dividing by V_{max} . In Fig. 4c, 5-HT_{3A}-injected oocytes were first tested in buffer containing 0.3 mM Ca²⁺ and then 10 mM Ca²⁺, as the 5-HT_{3A}-injected oocytes recovered very slowly from the 10 mM Ca²⁺ treatment. *mod-1*-injected oocytes were tested in both orders. The concentration of Ca²⁺ was varied in both the 5-HT and wash solutions. The 5-HT_{3A} traces were filtered digitally off-line at 5 Hz to reduce noise. In Fig. 4d and e, antagonists were present in both the 5-HT and wash solutions.

HEK293 cells were transiently transfected and whole-cell voltage-clamp recordings were done the next day as described³⁰. Output was filtered at 10 kHz and sampled at 20 kHz using a Digidata 1200 interface. The patch pipette was coated with Sylgard, and the tip was heat polished to a final tip resistance in bath solution of 0.5–2 M Ω . The HEK293 cell was lifted from the dish with the recording pipette, and 5-HT was applied from a pressure-controlled spritzer pipette in the presence of a continuously flowing bath solution. The internal (patch pipette) solution contained (in mM): 130 potassium aspartate, 10 KCl, 1 MgCl₂, 10 EGTA, 10 HEPES (pH 7.4). The external solution contained (in mM) 2 CaCl₂, 1 MgCl₂, 10 HEPES (pH 7.4), with varying proportions of sodium gluconate and NaCl, resulting in a final concentration of 140 mM of both Na⁺ cations and gluconate plus Cl⁻ anions. Data analysis was performed off-line using Clampfit software (Axon), and curve fitting was accomplished using Origin (Microcal) or KaledaGraph (Adelbeck) software.

Received 28 July; accepted 11 September 2000.

- Martin, G. R., Eglén, R. M., Hamblin, M. W., Hoyer, D. & Yocca, F. The structure and signalling properties of 5-HT receptors: an endless diversity? *Trends Pharmacol. Sci.* **19**, 2–4 (1998).
- Marić, A. V., Peterson, A. S., Brake, A. J., Myers, R. M. & Julius, D. Primary structure and functional expression of the 5HT₃ receptor, a serotonin-gated ion channel. *Science* **254**, 432–437 (1991).
- Sawin, E. R., Ranganathan, R. & Horvitz, H. R. *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. *Neuron* **26**, 619–631 (2000).
- Horvitz, H. R., Chalfie, M., Trent, C., Sulston, J. E. & Evans, P. D. Serotonin and octopamine in the nematode *Caenorhabditis elegans*. *Science* **216**, 1012–1014 (1982).
- The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
- Krause, M. & Hirsh, D. A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* **49**, 753–761 (1987).
- Liu, L. X. *et al.* High-throughput isolation of *Caenorhabditis elegans* deletion mutants. *Genome Res.* **9**, 859–867 (1999).
- Jansen, G., Hazendonk, E., Thijssen, K. L. & Plasterk, R. H. Reverse genetics by chemical mutagenesis in *Caenorhabditis elegans*. *Nature Genet.* **17**, 119–121 (1997).
- Sumikawa, K. & Gehle, V. M. Assembly of mutant subunits of the nicotinic acetylcholine receptor lacking the conserved disulfide loop structure. *J. Biol. Chem.* **267**, 6286–6290 (1992).
- Karlin, A. & Akabas, M. H. Toward a structural basis for the function of nicotinic acetylcholine receptors and their cousins. *Neuron* **15**, 1231–1244 (1995).
- Cooper, J. R., Bloom, F. E. & Roth, R. H. *The Biochemical Basis of Neuropsychopharmacology* (Oxford Univ. Press, New York, 1996).
- Hardie, R. C. A histamine-activated chloride channel involved in neurotransmission at a photoreceptor synapse. *Nature* **339**, 704–706 (1989).
- McClintock, T. S. & Ache, B. W. Histamine directly gates a chloride channel in lobster olfactory receptor neurons. *Proc. Natl Acad. Sci. USA* **86**, 8137–8141 (1989).
- Cully, D. F. *et al.* Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*. *Nature* **371**, 707–711 (1994).
- Adelsberger, H., Lepier, A. & Dudel, J. Activation of rat recombinant $\alpha_1\beta_2\gamma_{25}$ GABA_A receptor by the insecticide ivermectin. *Eur. J. Pharmacol.* **394**, 163–170 (2000).
- Krause, R. M. *et al.* Ivermectin: a positive allosteric effector of the α_7 neuronal nicotinic acetylcholine receptor. *Mol. Pharmacol.* **53**, 283–294 (1998).
- Davies, P. A. *et al.* The 5-HT_{3B} subunit is a major determinant of serotonin-receptor function. *Nature* **397**, 359–363 (1999).
- Downie, D. L. *et al.* Pharmacological characterization of the apparent splice variants of the murine 5-HT₃ R-A subunit expressed in *Xenopus laevis* oocytes. *Neuropharmacology* **33**, 473–482 (1994).
- Aizenberg, D. *et al.* Mianserin, a 5-HT_{2A/2C} and α_2 antagonist, in the treatment of sexual dysfunction induced by serotonin reuptake inhibitors. *Clin. Neuropharmacol.* **20**, 210–214 (1997).
- Granas, C. & Larhammar, D. Identification of an amino acid residue important for binding of methiothepin and sumatriptan to the human 5-HT_{1B} receptor. *Eur. J. Pharmacol.* **380**, 171–181 (1999).
- Weber, W. Ion currents of *Xenopus laevis* oocytes: state of the art. *Biochim. Biophys. Acta* **1421**, 213–233 (1999).
- Bormann, J., Hamill, O. P. & Sakmann, B. Mechanism of anion permeation through channels gated by glycine and gamma-aminobutyric acid in mouse cultured spinal neurones. *J. Physiol. (Lond.)* **385**, 243–286 (1987).
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).
- Shaham, S. & Horvitz, H. R. Developing *Caenorhabditis elegans* neurons may contain both cell-death protectors and killer activities. *Genes Dev.* **10**, 578–591 (1996).
- Lessmann, V. & Dietzel, I. D. Development of serotonin-induced ion currents in identified embryonic Retzius cells from the medicinal leech (*Hirudo medicinalis*). *J. Neurosci.* **11**, 800–809 (1991).
- Lessmann, V. & Dietzel, I. D. Two kinetically distinct 5-hydroxytryptamine-activated Cl⁻ conductances at Retzius P-cell synapses of the medicinal leech. *J. Neurosci.* **15**, 1496–1505 (1995).

27. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
28. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
29. Morrill, J. A. & Cannon, S. C. Effects of mutations causing hypokalaemic periodic paralysis on the skeletal muscle L-type Ca^{2+} channel expressed in *Xenopus laevis* oocytes. *J. Physiol. (Lond.)* **520**, 321–336 (1999).
30. Hayward, L. J., Brown, R. H. Jr & Cannon, S. C. Inactivation defects caused by myotonia-associated mutations in the sodium channel III-IV linker. *J. Gen. Physiol.* **107**, 559–576 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank N. Buttner, D. Omura and P. Reddien for suggestions concerning this manuscript; G. Moulder and R. Barstead for help in isolating the *mod-1(ok103)* allele; L. Liu and C. Johnson for sharing the *mod-1(nr2043)* allele before publication; and D. Julius for the 5-HT_{3A} cDNA clone. This work was supported by a grant from the United States Public Health Service (H.R.H.). R.R. is supported by a Howard Hughes Medical Institute predoctoral fellowship. H.R.H. is an Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to H.R.H. (e-mail: horvitz@mit.edu). The nucleotide sequence of the *mod-1* cDNA has been deposited at GenBank under the accession number AF303088.

Notch signalling and the synchronization of the somite segmentation clock

Yun-Jin Jiang*, Birgit L. Aerne†, Lucy Smithers*‡, Catherine Haddon*, David Ish-Horowicz† & Julian Lewis*

* Vertebrate Development Laboratory and † Developmental Genetics Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

In vertebrates with mutations in the Notch cell–cell communication pathway, segmentation fails: the boundaries demarcating somites, the segments of the embryonic body axis, are absent or irregular^{1–8}. This phenotype has prompted many investigations, but the role of Notch signalling in somitogenesis remains mysterious^{9–12}. Somite patterning is thought to be governed by a “clock-and-wavefront” mechanism¹³: a biochemical oscillator (the segmentation clock) operates in the cells of the presomitic mesoderm, the immature tissue from which the somites are sequentially produced, and a wavefront of maturation sweeps back through this tissue, arresting oscillation and initiating somite differentiation^{14,15}. Cells arrested in different phases of their cycle express different genes, defining the spatially periodic pattern of somites and controlling the physical process of segmentation^{1,16–19}. Notch signalling, one might think, must be necessary for oscillation, or to organize subsequent events that create the somite boundaries. Here we analyse a set of zebrafish mutants and arrive at a different interpretation: the essential function of Notch signalling in somite segmentation is to keep the oscillations of neighbouring presomitic mesoderm cells synchronized.

We have analysed the expression patterns of two zebrafish *Delta* homologues, *deltaC* and *deltaD*^{10,20,21}. These code for Notch ligands, and their patterns reveal both the existence of a segmentation clock in the zebrafish and its linkage to the Notch pathway. Expression of *deltaC* marks the posterior part of each formed somite; expression of *deltaD* marks the antero-medial part²⁰. Meanwhile, expression in the presomitic mesoderm (PSM) appears to be dynamic. In a batch

of sibling embryos all fixed at the same time, a variety of stripy PSM expression patterns of *deltaC* and *deltaD* are seen. This variation is most obvious for *deltaC*. The expression patterns can be arranged tentatively as phases of a cycle (Fig. 1), resembling the cycle of patterns of *c-hairy1* expression in the chick embryo¹⁴. Following the chick convention¹⁴, we divide the cycle of patterns into three intervals, labelled phases I, II and III. Embryos at different ages show the same range of patterns, with similar relative frequencies (315 embryos scored, stages from 1 to 10 somites), as expected for a repetitive cycle.

To prove that the different *deltaC* expression patterns represent consecutive states of a cyclic process, we analysed embryos developing in a temperature gradient that causes somitogenesis to go faster on one side of the body than on the other. Embryos at the 5–10-somite stage were placed on their sides in a chamber with a warm floor and a cool ceiling, giving a temperature gradient calculated to cause the rate of somitogenesis to be faster on the warm side of the body than on the cold side by 0.2–0.3 somites per hour (see Methods). We took batches of embryos that had been exposed to the temperature gradient for the same fixed time (1.5 h, corresponding to a left–right difference of 0.3–0.45 of a somite cycle) and compared the patterns of *deltaC* expression between the two sides (Fig. 2). Typically, when the cool side showed a phase I pattern (as defined in Fig. 1), the warm side showed a phase II pattern; when the cool side showed a phase II pattern, the warm side showed a phase III pattern; and when the cool side showed a phase III pattern, the warm side showed a phase I pattern (see Fig. 2 for details). The phases therefore succeed each other cyclically. Long periods in the temperature gradient caused more extreme differences between the two sides, and after 3.5 h embryos could be seen in which the warm side was once again in the same phase as the cool side, but with one extra somite (Fig. 2d).

The stripes of expression seen in the PSM correspond to repeated waves of *deltaC* expression, sweeping through a much more slowly moving cell population (see Supplementary Information). Like the *c-hairy1* waves in the chick¹⁴, these ‘kinematic waves’ are the sign of an intracellular oscillation that slows down, changes its amplitude and finally comes to a halt as cells progress from the posterior PSM to the anterior PSM and emerge into the region where somite boundaries finally become determined (see ref. 14 for analysis). The clock-and-wavefront mechanism therefore appears to be a fundamental feature of vertebrate somitogenesis, conserved from fish to birds and mammals.

Using the expression of *deltaC* as an indicator, we examined how oscillator behaviour is altered in the somitogenesis mutants *beamter* (*bea*), *deadly seven* (*des*), *after eight* (*aei*) and *mind bomb* (*mib*)^{22–24}. These all have a somite phenotype like that of Notch pathway mutants in the mouse: somite boundaries fail to form correctly, except for the first 5–8 (the most anterior), which appear normal. The *mib* mutation makes cells deaf to Notch signalling in other tissues^{23,25}, *aei* corresponds to *deltaD*²⁶ and *des* shows neurogenic abnormalities in the neural plate similar to those of *mib* but weaker (data not shown). Thus we can ascribe the somite phenotype in fish, as in mouse, to a defect in Notch signalling.

Figure 3 shows the *deltaC* expression pattern for each of the four mutants at the 10-somite stage. The expression pattern is the same in each case, and abnormal in four respects. First, all mutant embryos in a batch show the same pattern, with moderate levels of expression in the posterior part of the PSM, relatively low levels in the middle part, and high levels in the anterior part. Second, in the anterior PSM, the two strong stripes visible in most phases of the normal cycle are replaced by a single abnormally broad stripe. Third, levels of expression are abnormally variable from cell to cell; this is particularly obvious in the anterior PSM, where there is a random mixture of strongly and weakly expressing cells instead of the usual alternating stripes of very high and very low expression

‡ Present address: Gene Function and Regulation, Institute of Cancer Research, Chester Beatty Labs, Fulham Road, London SW3 6JB, UK.

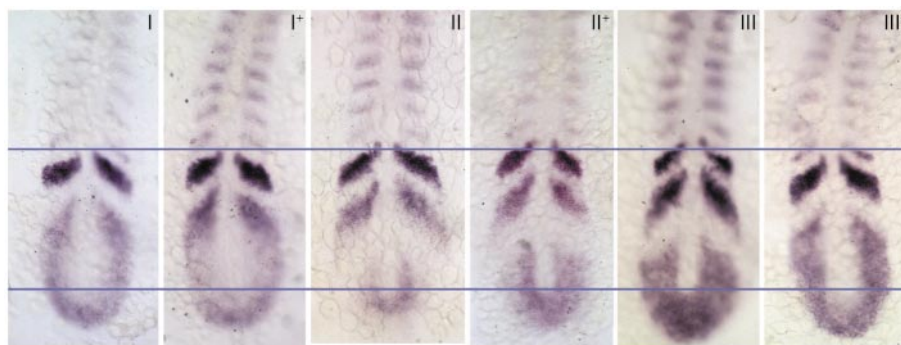


Figure 1 Expression of *deltaC* in normal sibling embryos fixed at about the 10-somite stage. The different patterns can be arranged in a sequence corresponding to a temporal oscillation in the PSM linked to somitogenesis. Phases of the oscillation cycle are

numbered I, II and III to match terminology for the *c-hairy1* cycle in the chick¹⁴. Intermediate phases are designated I+, II+ and III+. The posterior end of the notochord serves as a reference point for alignment of the embryos (horizontal guide lines).

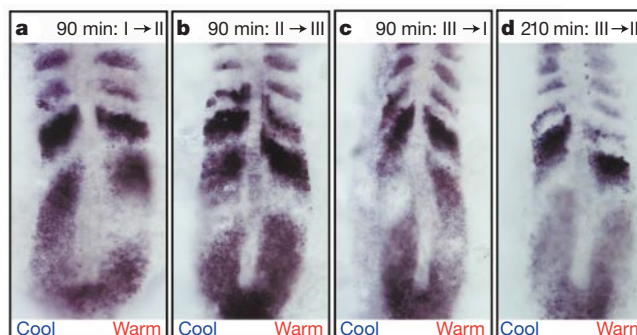


Figure 2 *deltaC* expression patterns in embryos cultured in a temperature gradient, to give faster development on the right side than on the left. **a–c**, After 1.5 h of culture. In each case, the warm side is in a phase one step later than the cool side, confirming the cyclic nature of the changes. (26/44 embryos showed a clear phase difference between the two sides. Numbers of cases of each type of cold-phase/warm-phase result were I/I,

6; I/II, 10; II/II, 4; II/III, 10; III/III, 8; III/I, 6.). **d**, After 3.5 h of culture. Both sides are now in the same phase (phase III), but one extra somite has formed on the warm side. (20/29 embryos showed an extra somite on the warm side, and in 11 of these cases the phases were the same on both sides.)

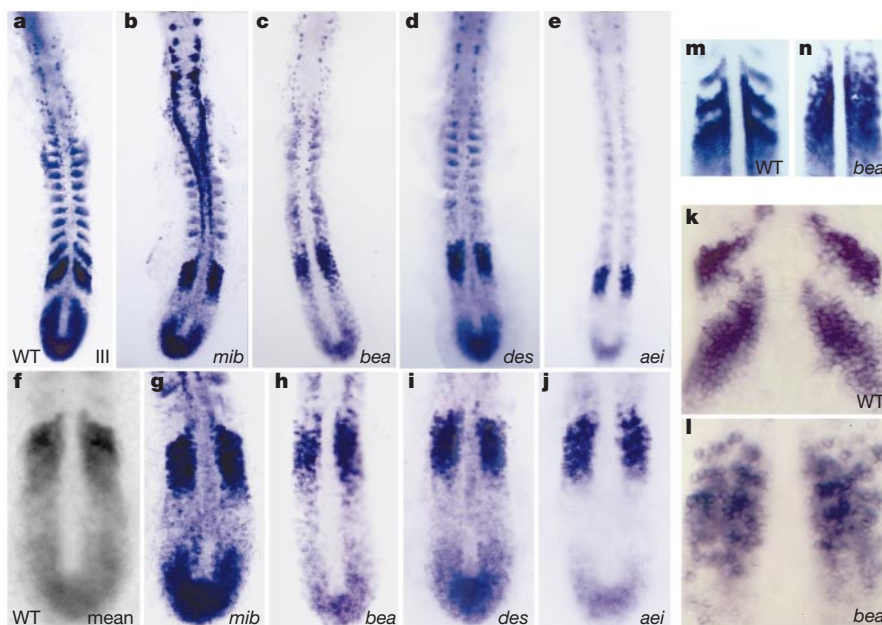


Figure 3 Expression of *deltaC* and *papc* in *mib*, *bea*, *des* and *aei* mutants. **a–l**, *deltaC* expression patterns in mutants and wild type, at the 10-somite stage. For each of the four mutations, every embryo shows a similar *deltaC* pattern in the PSM, with no sign of the organized oscillation seen in the wild type. **a–j**, Low magnification (**a–e**) and higher magnification (**f–j**) views showing the patterns in the PSM. The mutants are shown in **g–j**; **f** shows the time-averaged level of *deltaC* expression in the PSM of a normal embryo as a function of position. The average is computed using Photoshop to superimpose the

images in Fig. 1 (see Methods). This gives only a rough qualitative indication of true mean levels of expression, as the intensity of the *in situ* hybridization signal is a nonlinear function of the quantity of RNA. **k, l**, Details of anterior PSM in wild type (phase III) (**k**) and *bea* (**l**). **m, n**, *papc* expression at the 10-somite stage in wild type (**m**) and *bea* (**n**). In the region of nascent somites, the normal regular alternation of expressing (anterior) and non-expressing (posterior) cells is replaced in the mutant by a random mixture.

(Fig. 3g–j, k, l). Last, in the region that would normally correspond to formed somites, regular periodicity is lost, except for the most anterior 5–8 somites, where a faint periodic pattern of *deltaC* expression can still be seen.

At first glance, these data seem to show that the cells in the PSM are no longer cycling. Closer consideration suggests an alternative interpretation. The expression of *deltaC* is neither reduced to zero in the PSM, nor is it raised to a maximal level; instead, each region consists of a mixture of cells showing different levels of expression, with a local population average that is high where the wild-type oscillation maxima are high, low where they are low, and moderate where they are moderate (Fig. 3g–j, l). In other words, there is a loss of coordination between adjacent cells, and each region includes examples of all the different levels of *deltaC* expression that would occur in that region in the course of a normal cycle. The simplest interpretation is that the individual cells are continuing to oscillate in the mutants with an amplitude appropriate to the region of the PSM in which they lie, but that local cell–cell synchrony has been lost, creating a random mixture of cells in different phases of their individual oscillation cycles. If this is the case, a population average of the cells in any given region of the mutant PSM should be equivalent to a temporal average (over one complete somitogenesis cycle) of the states of the cells in that region in the wild type. A computed image of the wild-type temporal average (Fig. 3f) shows good qualitative agreement with the mutant *deltaC* patterns.

According to the desynchronization hypothesis, other genes that normally show oscillating expression in the PSM should have their expression patterns altered there in the same way that *deltaC* expression patterns are altered—that is, they should be locally phase-randomized. This matches the observations for *deltaD*, which, like *deltaC*, shows a grossly similar pattern in every

mutant embryo, with the usual stripes smeared out and levels of expression varying at random from cell to cell (data not shown). The expression pattern of the *hairy/E(spl)*-related gene *her1* (refs 26–28) shows the same type of disorder (Fig. 3 of ref. 24, and our own unpublished data).

Anterior to the PSM, in the region of nascent somites, *paraxial protocadherin* (*papc*)¹⁹ provides a marker of the pattern laid down as oscillation comes to a halt and cell characters become fixed. Normally in this region, bands of cells arrested in a state corresponding to the anterior part of a somite, which express *papc*, alternate with bands of cells arrested in a posterior somite state, which do not express *papc*. In the mutants, there is a random mixture of cells, some expressing *papc* and some not (Fig. 3m and n), just as the desynchronized clock-and-wavefront model predicts. Similar phenomena are seen for the anterior-somite marker *mesp-b*²⁸. (A quite different molecular abnormality occurs in a fifth somitogenesis mutant, *fused somites* (*fss*)²², superficially similar to *mib*, *bea*, *des* and *aei*, but with even its most anterior somites disrupted. In *fss*, *deltaC* shows normal cyclic patterns in the PSM, but expression of *papc* in nascent somites is lost, implying a defect in somite maturation; see Supplementary Information.)

Although we do not have the tools to show directly that individual PSM cells continue oscillating in *bea*, *des*, *aei* and *mib* mutants, the desynchronization hypothesis provides the simplest explanation of the observations. A more complex possibility would be that the mutations cause the cells to stop cycling regularly and instead to undergo random or chaotic temporal fluctuations with similar amplitude. It is also conceivable that the loss of coordination between neighbours could arise from disorderly cell movement, which could mix up cells in different oscillation phases. Measure-

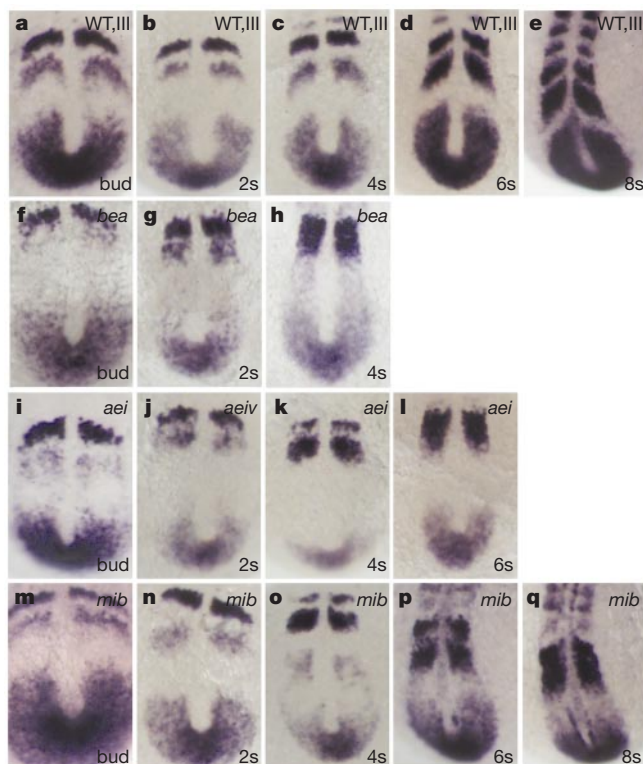


Figure 4 *deltaC* expression at early stages, from bud to 8 somites. **a–e**, Wild type. **f–h**, *bea*. **i–l**, *aei*. **m–q**, *mib*. Data for *des* (not shown) are similar to those for *aei*. Initially, the mutants, like their wild-type siblings, show multiple stripes of expression in the PSM; later, these signs of organized oscillation become blurred and are eventually lost. The onset of disorder in *deltaC* expression correlates with the onset of somite malformations (see Table 1).

Table 1 Onset of disturbance in *deltaC* expression in the PSM of somite mutants

Stage*	75%	90%	bud	2s	3s	4s	5s	6s	8s	10s
Mutant										
<i>bea</i>	–†	++	+	+	++	++		++		++
<i>des</i>			±	+	+		++‡			++
<i>aei</i>	–†	++	+	+	+	+	++	++		++
<i>mib</i>			–	–	–	–		±	++	++

Symbols: ++, 25% of embryos from a heterozygous mating show fusion of *deltaC* stripes in the PSM; +, 25% of embryos from a heterozygous mating show fusion or blurring, with some stripiness still evident; ±, some blurring seen, but in < 25% of embryos; –, all embryos show normal *deltaC* pattern. Each non-blank entry in the table represents a score based on examination of more than 40 offspring from a mating of heterozygotes or homozygotes.

*Stage terminology as in ref. 30; '75%' represents 75% epiboly; '1s', '2s', and so on, denote number of somites.

† Embryos from both heterozygous and homozygous matings.

‡ *des*^{h350} embryos.

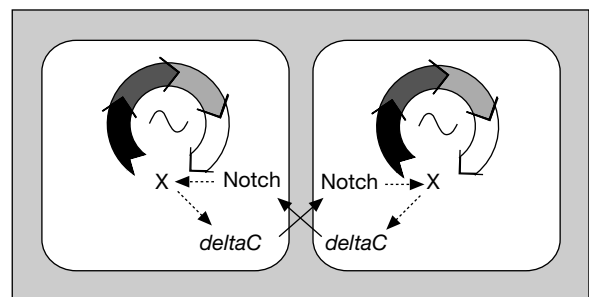


Figure 5 Signalling pathway proposed to keep oscillations in adjacent cells synchronized in the zebrafish. Each cell contains a biochemical oscillator capable of running independently of neighbouring cells. X denotes molecules of the oscillator mechanism (unknown). Cyclic changes in the activities of these molecules drive changes of *deltaC* expression in each cell and consequently changes of Notch activation in its neighbour. Notch activity exerts a relatively weak influence (dotted arrows) on the production of X. This interaction is sufficient to keep the oscillations in the adjacent cells synchronized.

ments from time-lapse movies (see Supplementary Information) make this latter interpretation improbable, however: in 10 min (1/3 of a somite cycle), a PSM cell in a *bea* embryo moves on average only 2.0 μm (r.m.s. displacement; $n = 290$ cells traced in 4 embryos) relative to its neighbours, and in *aei* only 4.9 μm ($n = 199$, 3 embryos), as compared with 2.8 μm ($n = 275$, 4 embryos) for wild-type controls. This is no more than about half a cell diameter for both mutant and wild type, and cannot account for blurring of a pattern of stripes on a scale of 50–100 μm .

The desynchronization hypothesis suggests a simple answer to the puzzle of why in mutants where all the rest of the somites are severely affected the first 5–8 somites are spared and develop normally. This sparing of anterior somites is a general feature of embryos with defects in the Notch pathway, seen not only in *bea*, *des*, *aei* and *mib*, but also, to varying degrees, in mice with mutations in *Notch1*, *Delta1*, *Delta3*, *RBPJK*, *Presenilin1* and *Lunatic fringe*^{1–8}. Yet wild-type embryos show no radical difference between the anterior and the posterior somites, either in morphology or in the pattern of expression of molecular markers of somite structure. It is unlikely that the anterior somites are spared in *mib*, *bea*, *des* and *aei* simply because of a higher level of redundancy in the control machinery that generates these somites, as double mutants show abnormalities no more severe than those of single mutants (ref. 24; and our own unpublished data). It is also unlikely that the sparing of anterior somites reflects persistence of maternal gene products, as the same phenomenon is seen in the progeny of homozygous *bea*, *des* and *aei* mutant parents. Suppose, however, that the PSM oscillator that controls somite patterning is set going synchronously in the PSM precursor cells at some time point early in gastrulation, both in the wild type and in the mutants. In normal development, the cells are kept locally synchronized in their subsequent oscillation cycles by cell–cell communication. If the cell–cell communication that maintains synchrony is defective, however, the cells will gradually drift out of synchrony until the lack of coordination causes somitogenesis to fail. As expected with this theory, the formation of the first few, relatively normal somites is foreshadowed in the PSM by a pattern of *deltaC* expression that is at first normal, with clear evidence of organized oscillation, and then gradually becomes blurred (Fig. 4 and Table 1). The periodicity of *deltaC* expression breaks down earliest in *bea*, which has the earliest onset of disturbed somite morphology, later in *des* and *aei*, which have later (more posterior) onset, and latest of all in *mib*, which has the latest (most posterior) onset. Progressive randomization of the oscillator phases of individual cells may explain not only the onset of local irregularities of somite formation, but also the loss of coordination of somitogenesis on the two sides of the embryo reported in mouse *Notch1* mutants².

We cannot exclude the possibility that some special, qualitatively different mechanism controls formation of the first few somites, but with our interpretation of the data there is no need to invoke any such complication: these somites can form by the same clock-and-wavefront mechanism as all the rest. Their normality in Notch pathway mutants implies that Notch signalling is not needed to keep individual cells oscillating, and that somite boundaries can form even when Notch signalling fails. The sole indispensable function of Notch-mediated cell–cell communication in somite segmentation, we argue, is to keep the oscillations of adjacent PSM cells synchronized.

For Notch signalling to maintain synchrony, the oscillator in each PSM cell must drive oscillatory activation of the Notch pathway in the neighbouring cells, and the level of Notch activation in these neighbours must influence their own intracellular oscillation (Fig. 5). A mechanism for the first effect is clear: in zebrafish, *deltaC* shows oscillating expression and codes for a Notch ligand; in chick and mouse, expression of the Notch modulator *Lunatic fringe* oscillates^{15,29}. As for the second effect, one can only speculate, since the mechanism of the PSM oscillator is unknown and Notch

signalling can regulate the expression of many different genes. It is a common observation, easily demonstrated by computer simulations, that weak interactions of almost any sort between oscillating systems with similar periods can entrain them to oscillate in synchrony. Thus in the PSM almost any type of influence of Notch activity on the intracellular oscillation should be capable of causing neighbouring cells to fall into the same rhythm. In zebrafish, at least, it seems that the normal influence of Notch signalling on the oscillator is indeed only weak: failure of signalling in the Notch pathway mutants causes no obvious change in the period of oscillation, as judged by the size of the anterior-most somites. However, drastic overactivation of the Notch pathway may perturb the oscillator severely¹⁰, and it would not be surprising if in some species a loss of Notch signalling were to affect not only the phasing of the oscillator, but also its amplitude and periodicity.

We have described a new type of function for Notch signalling: the synchronization of biochemical oscillations in adjacent cells. It remains to be seen whether similar phenomena occur in tissues other than presomitic mesoderm. □

Methods

Fish stocks, staining and imaging

We obtained zebrafish embryos by natural spawnings and maintained them normally at 28.5 °C in system water. We used the mutant alleles *bea*^{tw212b}, *des*^{dc225} (or sometimes *des*^{th35b}), *aei*^{tr233} and *mib*^{ms2b} (the white tail allele)^{22,23}. Embryos were staged as described³⁰. Digoxigenin-labelled RNA antisense probes were generated with a Stratagene RNA transcription kit. Enzymes for linearization and transcription for probe synthesis were as follows: *deltaC*^{20,21}, *XbaI* and *T7*; *deltaD*²⁵, *EcoRI* and *T7*; *papc*¹⁹, *Apal* and *T3*. Embryos were stained as whole mounts by *in situ* hybridization²⁰ and flat mounted and photographed in 100% glycerol or a mixture of benzyl benzoate and benzyl alcohol (2:1). Images were edited using Adobe Photoshop.

The time-averaged pattern of *deltaC* expression in Fig. 3f is computed by using Photoshop to superimpose the images in Fig. 1, each pattern being weighted according to its relative frequency in populations of normal sibling embryos (21% phase I + I⁺, 30% phase II + II⁺, 48% phase III + III⁺; 93 embryos scored at 9–10 somites).

Temperature gradient

For temperature-gradient experiments, we dechorionated the embryos and placed them on their left sides in 0.2% agarose gel in a chamber of depth 600 μm , with glass coverslips as the floor and roof (thicknesses 150 μm and 2 × 150 μm , respectively). The floor of the chamber rested on an aluminium hot block maintained at 37 °C, while the roof was covered with ice slush and maintained at 0 °C. Batches of embryos were kept in the differentially heated chamber for measured lengths of time, usually 90 min, to create discrepancies between the two sides of the body, and then fixed and processed for *in situ* hybridization with a *deltaC* probe. Because the thermal conductivities of glass and water (or agarose gel) are roughly equal, the set-up should ideally create a temperature gradient of 37/(600 + 3 × 150), that is, 0.035 °C μm^{-1} across the embryo; the actual gradient was somewhat less (~0.02 °C μm^{-1}) because of air gaps (~30 μm) between the chamber and the hot block, and between the two roof coverslips. The distance between the middles of the left-hand and right-hand slabs of somitic mesoderm is 100 μm , corresponding to a temperature difference of ~2–3 °C. The temperature coefficient for somitogenesis is ~0.11 somites per hour per °C (ref. 30). Thus, in our system the rate of development for the left-hand block of somitic mesoderm should be faster than for the right-hand block by ~0.2–0.3 somites per hour.

Received 20 July; accepted 5 September 2000.

- del Barco Barrantes, I. *et al.* Interaction between Notch signalling and Lunatic fringe during somite boundary formation in the mouse. *Curr. Biol.* **9**, 470–480 (1999).
- Conlon, R. A., Reaume, A. G. & Rossant, J. Notch1 is required for the coordinate segmentation of somites. *Development* **121**, 1533–1545 (1995).
- Evrard, Y. A., Lun, Y., Aulehla, A., Gan, L. & Johnson, R. L. *lunatic fringe* is an essential mediator of somite segmentation and patterning. *Nature* **394**, 377–381 (1998).
- Hrabé de Angelis, M., McIntyre, J. & Gossler, A. Maintenance of somite borders in mice requires the *Delta* homologue *Dll1*. *Nature* **386**, 717–721 (1997).
- Kusumi, K. *et al.* The mouse *pudgy* mutation disrupts *Delta* homologue *Dll3* and initiation of early somite boundaries. *Nature Genet.* **19**, 274–278 (1998).
- Oka, C. *et al.* Disruption of the mouse *RBP-Jk* gene results in early embryonic death. *Development* **121**, 3291–3301 (1995).
- Wong, P. C. *et al.* *Presenilin 1* is required for *Notch1* and *Dll1* expression in the paraxial mesoderm. *Nature* **387**, 288–292 (1997).
- Zhang, N. & Gridley, T. Defects in somite formation in *lunatic fringe*-deficient mice. *Nature* **394**, 374–377 (1998).
- Jen, W. C., Gawanitka, V., Pollet, N., Niehrs, C. & Kintner, C. Periodic repression of Notch pathway genes governs the segmentation of *Xenopus* embryos. *Genes Dev.* **13**, 1486–1499 (1999).
- Takke, C. & Campos-Ortega, J. A. *her1*, a zebrafish pair-rule like gene, acts downstream of Notch signalling to control somite development. *Development* **126**, 3005–3014 (1999).
- Pourquie, O. Notch around the clock. *Curr. Opin. Genet. Dev.* **9**, 559–565 (1999).

12. Jiang, Y.-J., Smithers, L. & Lewis, J. Vertebrate segmentation: the clock is linked to Notch signalling. *Curr. Biol.* **8**, R868–R871 (1998).
13. Cooke, J. & Zeeman, E. C. A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *J. Theoret. Biol.* **58**, 455–476 (1976).
14. Palmeirim, I., Henrique, D., Ish-Horowitz, D. & Pourquié, O. Avian *hairy* gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell* **91**, 639–648 (1997).
15. McGrew, M. J., Dale, J. K., Fraboulet, S. & Pourquié, O. The *lunatic fringe* gene is a target of the molecular clock linked to somite segmentation in avian embryos. *Curr. Biol.* **8**, 979–982 (1998).
16. Keynes, R. J. & Stern, C. D. Mechanisms of vertebrate segmentation. *Development* **103**, 413–429 (1988).
17. Saga, Y., Hata, N., Koseki, H. & Taketo, M. M. *Mesp2*: a novel mouse gene expressed in the presegmented mesoderm and essential for segmentation initiation. *Genes Dev.* **11**, 1827–1839 (1997).
18. Durbin, L. *et al.* Eph signaling is required for segmentation and differentiation of the somites. *Genes Dev.* **12**, 3096–3109 (1998).
19. Yamamoto, A. *et al.* Zebrafish *paraxial protocadherin* is a downstream target of *spadetail* involved in morphogenesis of gastrula mesoderm. *Development* **125**, 3389–3397 (1998).
20. Haddon, C. *et al.* Multiple *delta* genes and lateral inhibition in zebrafish primary neurogenesis. *Development* **125**, 359–370 (1998).
21. Smithers, L., Haddon, C., Jiang, Y.-J. & Lewis, J. Sequence and embryonic expression of *deltaC* in the zebrafish. *Mech. Dev.* **90**, 119–123 (2000).
22. van Eeden, F. J. *et al.* Mutations affecting somite formation and patterning in the zebrafish, *Danio rerio*. *Development* **123**, 153–164 (1996).
23. Jiang, Y.-J. *et al.* Mutations affecting neurogenesis and brain morphology in the zebrafish, *Danio rerio*. *Development* **123**, 205–216 (1996).
24. van Eeden, F. J., Holley, S. A., Haffter, P. & Nüsslein-Volhard, C. Zebrafish segmentation and pair-rule patterning. *Dev. Genet.* **23**, 65–76 (1998).
25. Haddon, C., Jiang, Y.-J., Smithers, L. & Lewis, J. Delta-Notch signalling and the patterning of sensory cell differentiation in the zebrafish ear: evidence from the mind bomb mutant. *Development* **125**, 4637–4644 (1998).
26. Holley, S. A., Geisler, R. & Nüsslein-Volhard, C. Control of *her1* expression during zebrafish somitogenesis by a *Delta*-dependent oscillator and an independent wave-front activity. *Genes Dev.* **14**, 1678–1690 (2000).
27. Müller, M., von Weizsäcker, E. & Campos-Ortega, J. A. Expression domains of a zebrafish homologue of the *Drosophila* pair-rule gene *hairy* correspond to primordia of alternating somites. *Development* **122**, 2071–2078 (1996).
28. Sawada, A. *et al.* Zebrafish *mesp* family genes, *mesp-a* and *mesp-b* are segmentally expressed in the presomitic mesoderm, and *mesp-b* confers the anterior identity to the developing somites. *Development* **127**, 1691–1702 (2000).
29. Aulehla, A. & Johnson, R. L. Dynamic expression of *lunatic fringe* suggests a link between Notch signaling and an autonomous cellular oscillator driving somite segmentation. *Dev. Biol.* **207**, 49–61 (1999).
30. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dynam.* **203**, 253–310 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We would like to thank J. Campos-Ortega and S. Holley for sharing data before publication; E. de Robertis for the *papc* probe; Q. Xu for advice on time-lapse filming; and H. McNeill for comments. The work was supported by the Imperial Cancer Research Fund and by an EMBO Fellowship to Y.-J.J.

Correspondence and requests for materials should be addressed to J.L. (e-mail: j.lewis@icrf.icnet.uk).

Binding of disease-associated prion protein to plasminogen

Michael B. Fischer*†, Christiane Roeckl*†, Petra Parizek*, Hans Peter Schwarz‡ & Adriano Aguzzi*

* Institute for Neuropathology, University Hospital of Zurich, CH-8091 Zürich, Switzerland

‡ Baxter Hyland Immuno, Industriestrasse 67, A-1221 Vienna, Austria

† These authors contributed equally to this study.

Transmissible spongiform encephalopathies are associated with accumulation of PrP^{Sc}, a conformer of a cellular protein called PrP^C. PrP^{Sc} is thought to replicate by imparting its conformation onto PrP^C (ref. 1), yet conformational discrimination between PrP^C and PrP^{Sc} has remained elusive. Because deposition of PrP^{Sc} alone is not enough to cause neuropathology², PrP^{Sc} probably damages the brain by interacting with other cellular constituents.

Here we find activities in human and mouse blood which bind PrP^{Sc} and prion infectivity, but not PrP^C. We identify plasminogen, a pro-protease implicated in neuronal excitotoxicity^{3,4}, as a PrP^{Sc}-binding protein. Binding is abolished if the conformation of PrP^{Sc} is disrupted by 6M urea or guanidine. The isolated lysine binding site 1 of plasminogen (kringles I–III) retains this binding activity, and binding can be competed for with lysine. Therefore, plasminogen represents the first endogenous factor discriminating between normal and pathological prion protein. This unexpected property may be exploited for diagnostic purposes.

An affinity assay was established for proteins that bind scrapie-associated PrP^{Sc} or its protease-resistant core PrP^{27–30}, but do not bind PrP^C. Paramagnetic beads were covalently coupled to serum proteins, and incubated with preparations containing PrP^C, PrP^{Sc} or PrP^{27–30}. Bound proteins were eluted and assayed by western blot with anti-PrP antibodies, or by bioassay with scrapie-susceptible mice.

Immobilized mouse serum proteins (1 mg ml^{−1}) precipitated PrP^{27–30} but neither native PrP^{Sc} nor PrP^C, whereas murine or human albumin did not precipitate any forms of PrP (Fig. 1a). Addition of non-infectious brain homogenate from C57BL/6 or *Prnp*^{0/0} mice⁵ digested with proteinase K (PK) allowed beads to bind PrP^{Sc} in addition to PrP^{27–30}; yet addition of phenylmethylsulphonyl fluoride (PMSF)-inactivated PK had no influence on binding activity (Fig. 1b). When coupled to beads at 0.1 mg ml^{−1}, serum no longer showed binding activity (data not shown), whereas monoclonal antibody 6H4 (ref. 6) bound PrP^C, PrP^{Sc} and PrP^{27–30} efficiently. Pre-immune mouse IgG and albumin never precipitated any form of PrP (Fig. 1c). These phenomena suggested that serum contains factors (collectively termed PrP_B) that interact specifically with the protease-resistant core of disease-associated prion protein. PrP_B was also found in serum of humans, sheep, cows and terminally scrapie-sick CD-1 mice (data not shown).

Why was PrP^{27–30} but not full-length PrP^{Sc} bound? If native PrP^{Sc} in the brains of sick mice were saturated with PrP_B, partial proteolysis may expose binding sites, or create PrP_B binding sites *de novo* on PrP^{Sc}. However, PrP^{Sc} binding was restored by addition of PK-digested non-infectious brain homogenate, indicating that neither hypothesis fully accounts for the binding properties of serum, and suggesting that unidentified proteolytic fragments liberated from brain extracts by PK digestion contribute to binding.

The template-directed refolding hypothesis⁷ predicts that PrP^C and PrP^{Sc} form heterodimers during prion replication. Therefore PrP_B may be identical with PrP^C. However, serum of *Prnp*^{0/0} mice (coupled to beads at 1 mg ml^{−1}) bound PrP^{27–30} (but neither PrP^{Sc} nor PrP^C) similarly to wild-type serum, indicating that PrP^C does not contribute to the binding activity of the serum (Fig. 1d).

We then attempted to enrich PrP_B from mouse serum by differential precipitation with ammonium sulphate. Each fraction was coupled to beads in the presence of 1 mg ml^{−1} protein. Most PrP_B precipitated at ammonium sulphate saturation below 50% (Fig. 2a). Although rabbit immunoglobulins raised against total mouse serum did not contain PrP_B (data not shown), they efficiently bound PrP^{27–30} upon pre-incubation with mouse serum or with proteins precipitating between 25% and 50% ammonium sulphate saturation; pre-incubation with 75–100% ammonium sulphate precipitates did not show any PrP_B activity (Fig. 2b). Similar results were obtained with human serum (data not shown). We conclude that PrP_B does not require covalent cross-linking to a solid surface to exert its activity, as PrP^{Sc} is precipitated in an indirect binding (sandwich) assay.

To investigate whether PrP^{27–30} binding activity correlates with immobilization of prion infectivity, an aliquot of bead suspension was implanted intracerebrally (i.c.) into *tga20* transgenic mice⁸. Upon exposure to infectious brain homogenate, beads with PrP_B activity induced scrapie in all inoculated animals: calculated prion

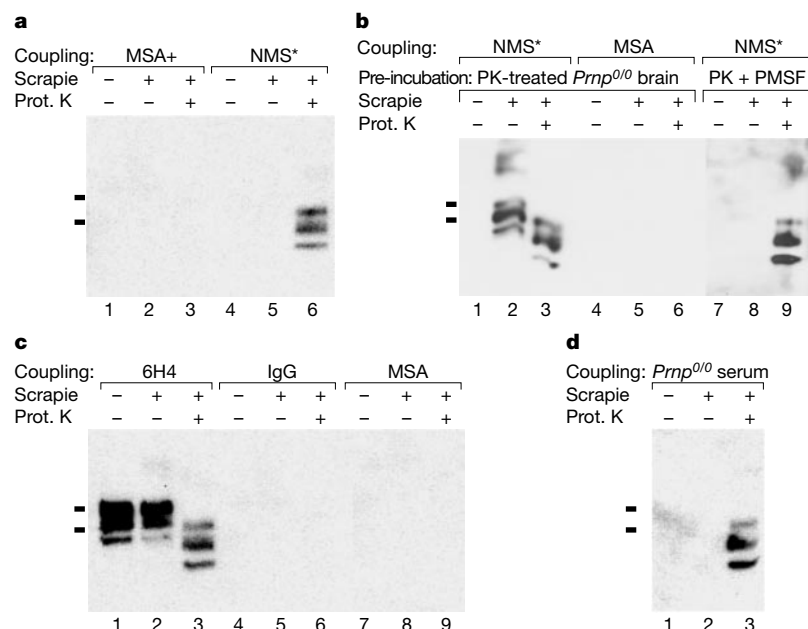


Figure 1 Western blot analysis of serum factors that interact with disease-associated prion protein. *Scrapie*⁻, brain homogenates of uninfected mice; *Scrapie*⁺, brain homogenates of terminally scrapie-sick CD-1 mice. After proteolytic digest (Prot. K+), PrP^C is degraded, while digestion of PrP^{Sc} results in a protease-resistant core of 27–30 kilodaltons termed PrP^{27–30}. This leads to a characteristic increase in electrophoretic mobility of the three bands diagnostic for PrP. **a**, When coupled to beads in the presence of 1 mg ml⁻¹ protein, mouse serum albumin (MSA*) did not show any affinity to any form of PrP (lanes 1–3) whereas normal mouse serum (NMS*) precipitated PrP^{27–30} (lanes 4–6) similarly to monoclonal antibody 6H4. Asterisks denote samples

coupled at a concentration of 1 mg ml⁻¹; all other samples were coupled at 0.1 mg ml⁻¹. **b**, After addition of PK-digested brain homogenate from *Prnp*^{0/0} mice, beads coupled to NMS bound PrP^{Sc} in addition to PrP^{27–30} (lanes 1–3), whereas beads coupled to albumin did not bind any PrP form (lane 4–6). The addition of inactive PK had no influence on binding activity (lane 7–9). **c**, 6H4 (monoclonal mouse antibody against PrP) coupled to beads bound all forms of PrP (lanes 1–3), whereas polyclonal mouse IgG (lanes 4–6) or MSA (lanes 7–9) did not bind any PrP (all beads were coupled in the presence of 0.1 mg ml⁻¹ protein). **d**, Serum of *Prnp*^{0/0} mice contains PrP_B similarly to wild-type serum. Molecular mass markers (top to bottom): 35 kilodaltons, 28 kilodaltons.

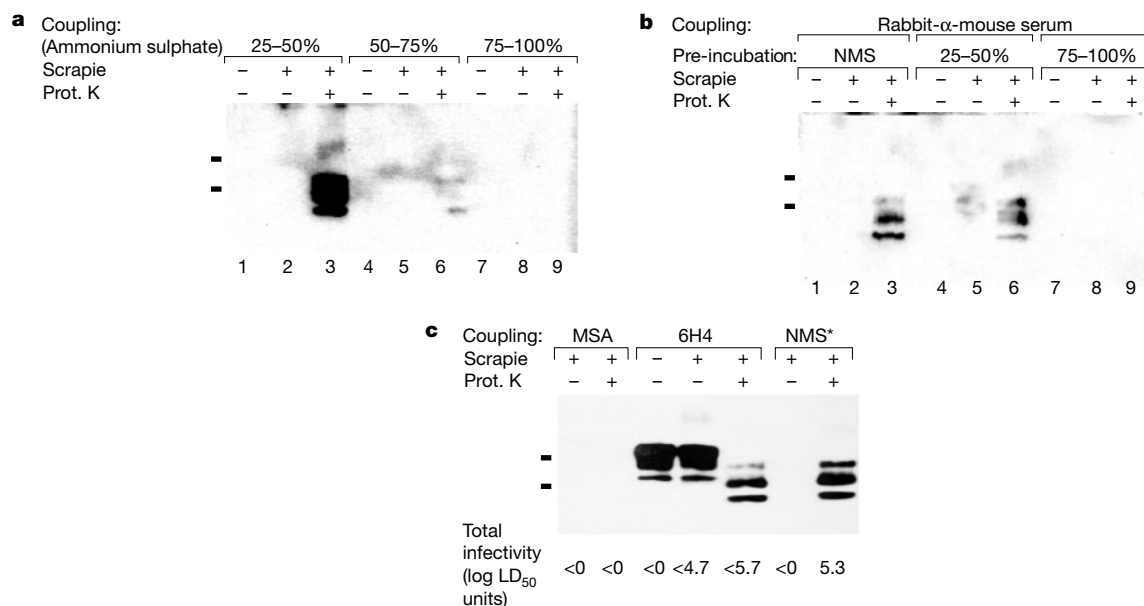


Figure 2 Partial purification of PrP_B. **a**, Serum was fractionated into three ammonium sulphate cuts, each of which was assayed for PrP_B activity. PrP_B was precipitated at 25–50% ammonium sulphate saturation (lanes 1–3). Serum proteins precipitating at an ammonium sulphate saturation of 50–75% showed little (lanes 4–6), and those above 75% (lanes 7–9) showed no PrP_B activity. **b**, Purified rabbit immunoglobulins raised against total mouse serum captured PrP_B activity upon pre-incubation with full serum (lanes 1–3) or with a 25–50% ammonium sulphate cut thereof (lanes 4–6). Pre-incubation with the 75–100% ammonium sulphate cut did not show PrP_B activity (lanes

7–9). **c**, Western blot analysis of eluates from beads coupled to mouse serum albumin, antibody 6H4, or normal mouse serum, and incubated with brain extracts as indicated. Before elution, an aliquot of paramagnetic beads was inoculated into *tga20* mice, which overexpress *Prnp*, for prion titre determinations. Infectivity titres were calculated using average incubation times as described, reported underneath each lane. Positive western blot signals match positive bioassay results; therefore PrP_B precipitates prion infectivity in addition to PrP^{27–30}.

infectivity titres were at least 10^5 times those of beads lacking PrP_B (Fig. 2c). Therefore PrP_B binds not only PrP²⁷⁻³⁰, but also prion infectivity.

Next, 61 fractions of human plasma were obtained by chromatography and differential precipitation, coupled to beads (protein concentration, 0.1 mg ml^{-1}) and tested for binding activity. Antibody 6H4 and mouse IgG, coupled under the same conditions, were used for controls. Thirteen fractions, as well as purified plasminogen, fibrinogen, antithrombin III and factor IX, bound PrP²⁷⁻³⁰. However, only plasminogen and fibrinogen bound non-digested PrP^{Sc} (Fig. 3a). Of the 41 fractions that tested negative, 6 contained defined proteins (albumin, factor XIII, thrombin, α_2 -macroglobulin, α_1 -antitrypsin and haptoglobin).

Plasminogen was the only protein binding full-length PrP^{Sc} in the absence of Ca^{2+} ions: in the presence of 10 mM EDTA, fibrinogen bound PrP²⁷⁻³⁰ but not PrP^{Sc} (Fig. 3b); whole serum also bound PrP²⁷⁻³⁰ in the presence of 10 mM EDTA or of 10 mM sodium citrate (data not shown). However, as serum did not bind PrP^{Sc}

from non-digested brain homogenate (Fig. 1a, lane 5), the interaction of plasminogen with disease-associated PrP is distinct from that of serum. All further experiments focused on plasminogen, which also precipitated human PrP^{CJD} from brain homogenates of patients with Creutzfeldt–Jakob disease, similarly to mouse PrP^{Sc} (data not shown).

Because plasminogen bound PrP^{Sc} but not PrP^C, interaction may be conformation-specific. When the assay was carried out under increasing urea concentrations, PrP^{Sc} became protease-sensitive in the presence of 6 M urea, indicating that PrP^{Sc} was largely denatured, and plasminogen no longer bound PrP^{Sc} or PrP²⁷⁻³⁰ (Fig. 3c). Therefore interaction requires the native conformation of either of the two partners, or of both. To assess whether binding of native plasminogen depends on the conformation of PrP^{Sc}, we denatured PrP^{Sc} by preincubation with 6 M guanidine thiocyanate. After dilution of guanidine to a concentration that was shown not to disturb the assay when present during the entire procedure, no monomeric PrP^{Sc} was recovered by precipitation with native

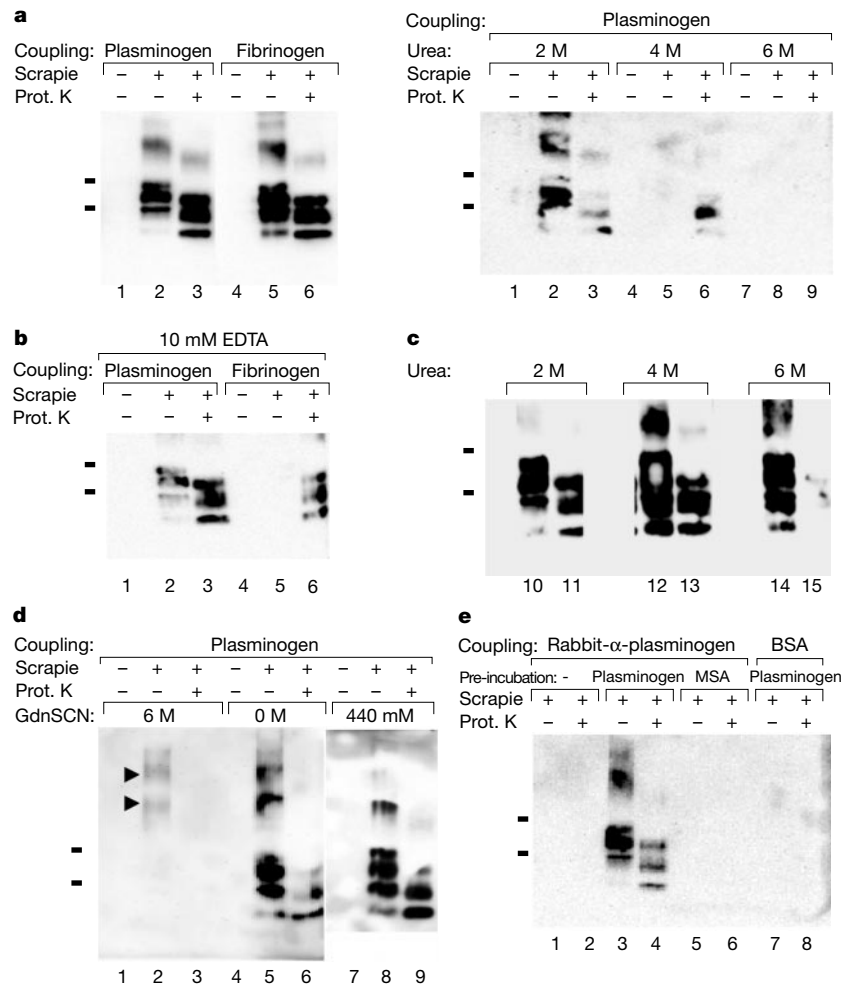


Figure 3 Characterization of the PrP^{Sc} binding activity of plasminogen. **a**, Purified plasminogen (lanes 1–3) and fibrinogen (lanes 4–6) fulfil the properties of PrP_B. **b**, The PrP^{Sc} binding activity of plasminogen is not abolished by chelation of Ca^{2+} ions, while fibrinogen only binds PrP²⁷⁻³⁰ (but not full-length PrP^{Sc}) in the presence of ethylenediaminetetraacetic acid (EDTA). **c**, Brain homogenates were incubated with increasing urea concentrations. PrP_B activity of plasminogen is detectable in the presence of 4 M urea, but is abolished by 6 M urea (lanes 7–9). An aliquot of each urea-exposed sample was PK-digested and analysed by western blotting (lanes 10–15): at 6 M urea PrP^{Sc} became PK-sensitive as evidenced by the lack of PrP²⁷⁻³⁰ detection (lanes 14–15). **d**, Native plasminogen does not interact with denatured PrP^{Sc}. Brain homogenates were denatured by pre-incubation with 6 M guanidine thiocyanate for 18 h. Binding assays

were performed after dilution of guanidine to 440 mM (lanes 1–3): this concentration did not impair plasminogen binding to PrP^{Sc} if brain homogenate had not been denatured (lanes 7–9), nor did incubation without guanidine thiocyanate for 18 h at room temperature (lanes 4–6). No monomeric PrP^{Sc} could be recovered by precipitation with native plasminogen. But a small amount of oligomeric PrP^{Sc} appeared to resist denaturation, or to renature upon dilution, and was still precipitated by plasminogen (arrowheads). **e**, Beads coated with anti-plasminogen antibodies did not show PrP_B activity (lanes 1 and 2). Upon pre-incubation with $100 \mu\text{g ml}^{-1}$ plasminogen (lanes 3 and 4), but not with albumin (lanes 5 and 6), PrP^{Sc} and PrP²⁷⁻³⁰ were bound. Preincubation of BSA-coupled beads with plasminogen did not result in PrP_B activity (lanes 7 and 8).

plasminogen (Fig. 3d). Conversely, plasminogen retained full PrP^{Sc} binding activity in 1 M NaCl (data not shown), suggesting that the interaction is unlikely to be electrostatic.

Covalent solid-state linking might partially denature proteins and expose hydrophobic patches to the solvent, which could bind aggregated PrP^{Sc} or PrP²⁷⁻³⁰ non-specifically. But this mechanism is unlikely to be significant, because magnetic beads coated with antibodies against plasminogen precipitated PrP^{Sc} when preincubated with plasminogen, yet not upon preincubation with albumin, nor when albumin-coupled beads were preincubated with plasminogen (Fig. 3e).

The correlation between immobilized PrP^{Sc} and prion infectivity was assessed by i.c. implantation of beads into *tga20* transgenic mice⁸ after exposure to infectious brain homogenate. Beads coupled to plasminogen induced scrapie in all inoculated animals (Fig. 4a): infectivity titres were up to 10⁵ times those of beads coupled to albumin, and those of plasminogen-coupled beads that were not incubated with brain homogenate.

Plasminogen kringle domains bind to lysine residues of interacting proteins. Preincubation of plasminogen-linked beads with 0.1 M lysine prevented binding to PrP^{Sc} (Fig. 4b). Inhibition was specific, because pre-incubation with 0.1 M arginine had no effect, and exposure of 6H4-linked beads to the same concentration of lysine did not prevent immunoprecipitation of PrP^{Sc} and of PrP^C. Moreover, purified plasminogen lysine binding site I (which contains the first three kringles of the plasmin A-chain) exhibited the same PrP^{Sc} binding activity as plasminogen (Fig. 4c). Therefore lysine residues of PrP^{Sc} are most probably responsible for its specific

interaction with plasminogen. Mouse PrP^C contains three lysines spaced 0.7–1 nm from each other⁹ in its structured domain, and seven further lysines in its flexible tail¹⁰. The octa repeat region of PrP appears to contribute to binding, since plasminogen precipitated PrP²⁷⁻³⁰ but not PrP^{Sc} from scrapie-infected {*tgd11;Prnp*^{0/0}} mice⁸ which express exclusively amino-terminally truncated (Δ 32–80) PrP (data not shown).

Codon 219 is polymorphic (Lys/Glu) in humans; *Prnp*-Lys 219 protects against prion disease¹¹ and chimaeric mouse/human *Prnp*-Lys 219 influences dominant-negatively PrP^{Sc} production¹². Lys 219 was suggested to interfere with binding of PrP^{Sc} to an unidentified chaperone provisionally named "protein X"¹²⁻¹⁴. Our data indicate that this interference may be effected by plasminogen.

Despite early claims⁶, no available immunoreagents reliably discriminate between PrP^{Sc} and PrP^C. Therefore, plasminogen is the first naturally occurring protein which selectively binds PrP^{Sc}. This interaction may have pathogenetic significance: perhaps PrP^{Sc} leads to local impairment of plasmin proteolysis, which is important for synaptic remodelling^{15,16}. Plasmin is restricted to rafts of cultured hippocampal neurons¹⁷, to which PrP^C also localizes¹⁸. If PrP^{Sc} interacts with further kringle proteins with trophic properties, their sequestration may play a role in pathogenesis of prion diseases.

The astonishing efficiency with which solid phase-bound plasminogen precipitates PrP^{Sc} might be exploited diagnostically. Moreover, blood is likely to be a carrier of bovine spongiform encephalopathy prions¹⁹. If plasminogen depletion reduces the infectious load of prion-spiked plasma, it might be useful for improving prion removal from blood-derived biological products. Finally, the structural basis for the interaction described here is likely to provide insights into the biology of prion diseases and into the physical nature of the infectious agent. □

Methods

Coupling to magnetic beads

100 μ g or 1 mg (see text) protein was diluted in 1 ml coupling buffer (0.1 M borate, pH 9.5). Tosyl-activated paramagnetic Dynabeads M-280 (Dyna, 1 ml) were washed twice with phosphate-buffered saline (PBS). Beads were rocked in coupling buffer containing proteins of interest for 24 h at 37 °C, and washed twice in PBS containing 0.1% w/v bovine serum albumin (BSA) for 5 min at room temperature (RT), once in blocking buffer (0.2 M Tris, pH 8.5, 0.1% BSA) for 4 h at 37 °C, once in PBS/BSA for 5 min at RT, once in 1% Tween-20 (Fluka, Buchs) for 10 min at RT, and once in PBS/BSA for 5 min at RT. Beads were stored in PBS/BSA containing 0.02% sodium azide.

PrP affinity assay

PBS containing 0.5% NP-40 and 0.5% sodium deoxycholate (DOC) was added to minced brain tissue (10% w/v) from uninfected mice or from mice infected with RML scrapie prions, 5th passage². Tissue was homogenized and centrifuged for 30 min at 500g at 4 °C. Supernatant was again centrifuged for 30 min at 500g at 4 °C. The second supernatant was adjusted to a protein concentration of 5 mg ml⁻¹ (in PBS, 0.5% NP-40, 0.5% DOC). Before analysis, 3 μ l PK (Roche Molecular Biochemicals, 215 μ g ml⁻¹), or 3 μ l PBS (as indicated), was added to 10 μ l brain homogenate (5 mg ml⁻¹) to a final concentration of 50 μ g ml⁻¹. All samples were incubated for 30 min at 37 °C. After addition of 3 μ l of 100 mM PMSF (Merck, Dietikon) and 8 μ l loading buffer (150 mM Tris-HCl, pH 6.8, 6% sodium dodecyl sulphate (SDS), 0.03% bromophenol blue, 30% glycerol), preparations were heated to 95 °C for 5 min. Samples were run immediately, or stored at –20 °C in which case they were heated again for 30 s at 95 °C before loading on gels.

For each protein to be tested for PrP^B activity, three samples were prepared. Sample I was produced by adding 1 ml assay buffer (PBS, 3% NP-40, 3% Tween-20) to 10 μ l of non-infected brain homogenate. Samples II and III were derived by adding 1 ml assay buffer to 10 μ l of infected brain homogenate. 50 μ l PK (1 mg ml⁻¹) was added to sample III to a final concentration of 50 μ g ml⁻¹. All samples were incubated for 30 min at 37 °C, and 50 μ l of 100 mM PMSF was added. For urea denaturation experiments, samples I–III (after PK digestion where indicated) were incubated for 2 h with 1 ml assay buffer containing urea at various concentrations. Urea concentration was kept constant throughout the assay, including washing steps. Instead, for denaturation experiments with guanidine thiocyanate, samples I–III (after PK digestion where indicated) were incubated with 6 M guanidine thiocyanate (18 h, RT, total volume 80 μ l) and, before addition of beads, diluted with 1 ml assay buffer to a final concentration of 440 mM guanidine thiocyanate. Beads (50 μ l) coupled to the proteins of interest were added, and incubated under continuous mixing for 1.5 h at RT.

For indirect precipitations, beads were preincubated before exposure to samples I–III with the protein of interest in assay buffer for 1.5 h at RT under continuous mixing, and washed three times with 1 ml washing buffer (PBS, 2% NP-40, 2% Tween 20). Where

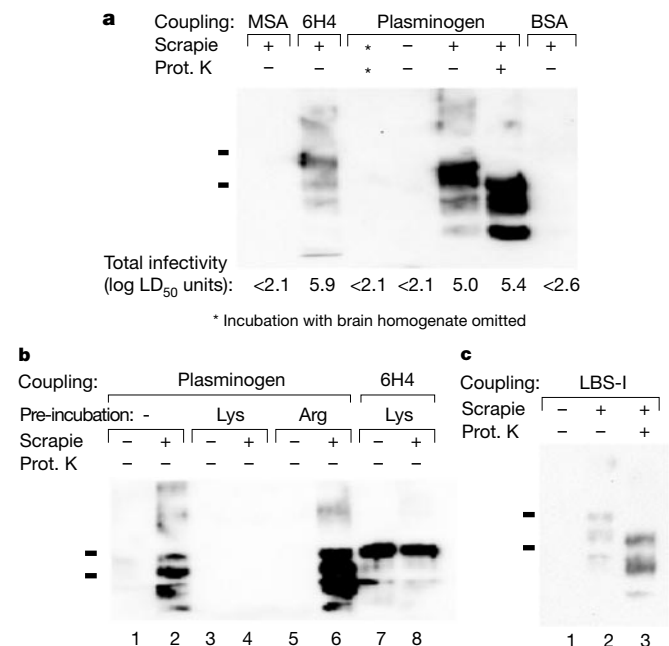


Figure 4 The mechanism by which plasminogen binds PrP^{Sc} and prion infectivity. **a**, Eluates from beads were subjected to western blot analysis and prion titre determination by bioassay with *tga20* mice: infectivity titres associated with the samples are reported underneath each lane. Average incubation times were used to calculate infectivity titres as described. All beads that bound PrP²⁷⁻³⁰ or PrP^{Sc} induced scrapie in indicator mice (infectivity titres were calculated by the incubation time method), indicating that plasminogen immobilizes prion infectivity in addition to PrP^{Sc}. **b**, PrP^B activity of plasminogen-linked beads (lanes 1 and 2) was abrogated in the presence of 100 mM lysine (lanes 3 and 4); binding was unaffected by 100 mM arginine (lanes 5 and 6). Binding of 6H4-linked beads to PrP^{Sc} was not affected by 100 mM lysine (lanes 7 and 8). **c**, Isolated lysine binding site I (kringles I–III, LBS-I) of plasminogen possess full binding activity for both PrP^{Sc} and PrP²⁷⁻³⁰.

indicated, beads were preincubated in assay buffer containing 0.1 M lysine or 0.1 M arginine (pH 7.3, 1 h, RT) before addition of brain homogenate. Washing buffer also contained 0.1 M lysine or arginine. After each washing step, beads were collected using a magnetic particle concentrator (MPC, Dynal).

Beads were washed three times with 1 ml washing buffer (containing 2–6 M urea where appropriate) and once with 1 ml PBS by vortexing for 15 s at RT and collecting with the MPC. Finally, beads were sedimented by centrifugation (no other centrifugation steps were performed), supernatant was discarded, and loading buffer (24 µl) was added. The preparation was heated to 95 °C for 5 min. Samples were run immediately or stored at –20 °C, in which case they were reheated before loading.

Western blot analysis

Bead eluates (24 µl) or brain homogenates (12 µl) were run on SDS-PAGE (5% stacking and 12% resolving) and blotted on nitrocellulose membranes (Schleicher & Schuell). Membranes were blocked with 5% (w/v) nonfat dry milk (NDM) in Tris-buffered saline–Tween (TBS–T) at RT for 30 min, incubated with 12.5 ml TBS–T/1% NDM and 5 µg antibody 6H4 (Prionics, Zürich) at RT under agitation for 1 h, washed three times for 10 min in TBS–T, and incubated with 12.5 ml TBS–T containing 1% NDM and rabbit anti-mouse IgG1-horseradish peroxidase (Zymed, 1:10,000) at RT under agitation for 30 min. Membranes were washed three times for 15 min in TBS–T, and developed using enhanced chemiluminescence (ECL) detection reagents. Signals were recorded on film and/or quantified using a Kodak ImageStation.

Infectivity bioassay

Groups of 3–5 *tga20* transgenic mice were inoculated i.c. with 20 µl PBS containing 0.2% of the beads used for precipitation. Implantation of non-infectious beads did not induce clinical signs; histological examination revealed mild perifocal gliosis but no spongiosis. Incubation time until development of terminal scrapie sickness was determined, and total infectivity titres in beads used for precipitation were calculated²⁰ using the relationship $y = 12.63 - 0.088x$, where y is the ID₅₀ and x is incubation time (days) to terminal disease².

Received 12 July; accepted 19 October 2000.

- Prusiner, S. B. Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136–144 (1982).
- Brandner, S. *et al.* Normal host prion protein necessary for scrapie-induced neurotoxicity. *Nature* **379**, 339–343 (1996).
- Chen, Z. L. & Strickland, S. Neuronal death in the hippocampus is promoted by plasmin-catalyzed degradation of laminin. *Cell* **91**, 917–925 (1997).
- Tsirka, S. E., Rogove, A. D. & Strickland, S. Neuronal cell death and tPA. *Nature* **384**, 123–124 (1996).
- Büeler, H. R. *et al.* Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein. *Nature* **356**, 577–582 (1992).
- Korth, C. *et al.* Prion (PrP^{Sc})-specific epitope defined by a monoclonal antibody. *Nature* **390**, 74–77 (1997).
- Prusiner, S. B. & Scott, M. R. Genetics of prions. *Annu. Rev. Genet.* **31**, 139–175 (1997).
- Fischer, M. *et al.* Prion protein (PrP) with amino-proximal deletions restoring susceptibility of PrP knockout mice to scrapie. *EMBO J.* **15**, 1255–1264 (1996).
- Riek, R. *et al.* NMR structure of the mouse prion protein domain Prp(121–231). *Nature* **382**, 180–182 (1996).
- Riek, R., Hornemann, S., Wider, G., Glockshuber, R. & Wüthrich, K. NMR characterization of the full-length recombinant murine prion protein, mPrP(23–231). *FEBS Lett.* **413**, 282–288 (1997).
- Shibuya, S., Higuchi, J., Shin, R. W., Tateishi, J. & Kitamoto, T. Protective prion protein polymorphisms against sporadic Creutzfeldt–Jakob disease. *Lancet* **351**, 419 (1998).
- Zulianiello, L. *et al.* Dominant-negative inhibition of prion formation diminished by deletion mutagenesis of the prion protein. *J. Virol.* **74**, 4351–4360 (2000).
- Telling, G. C. *et al.* Prion propagation in mice expressing human and chimeric PrP transgenes implicates the interaction of cellular PrP with another protein. *Cell* **83**, 79–90 (1995).
- Kaneko, K. *et al.* Evidence for protein X binding to a discontinuous epitope on the cellular prion protein during scrapie prion propagation. *Proc. Natl Acad. Sci. USA* **94**, 10069–10074 (1997).
- Madani, R. *et al.* Enhanced hippocampal long-term potentiation and learning by increased neuronal expression of tissue-type plasminogen activator in transgenic mice. *EMBO J.* **18**, 3007–3012 (1999).
- Baranes, D. *et al.* Tissue plasminogen activator contributes to the late phase of LTP and to synaptic growth in the hippocampal mossy fiber pathway. *Neuron* **21**, 813–825 (1998).
- Ledesma, M. D. *et al.* Brain plasmin influences APP alpha-cleavage and Aβ degradation and is reduced in Alzheimer's disease brains. *EMBO Rep.* (in the press).
- Naslavsky, N., Stein, R., Yanai, A., Friedlander, G. & Taraboulos, A. Characterization of detergent-insoluble complexes containing the cellular prion protein and its scrapie isoform. *J. Biol. Chem.* **272**, 6324–6331 (1997).
- Houston, F., Foster, J. D., Chong, A., Hunter, N. & Bostock, C. J. Transmission of BSE by blood transfusion in sheep. *Lancet* **356**, 999–1000 (2000).
- Prusiner, S. B. *et al.* Measurement of the scrapie agent using an incubation time interval assay. *Ann. Neurol.* **11**, 353–358 (1982).

Acknowledgements

We are grateful to D. Voelkel for providing fractionated human plasma, to P. Sonderegger, C. Weissmann, T. Lührs, and W. Schaffner for critical advice, and to M. Maissen and R. Hardegger for experimental help. This study was supported by grants from the European Union (Bundesamt für Bildung und Wissenschaft), the Swiss National Research Programs to A.A. and by a grant-in-aid of the SigmaXi foundation to M.B.F.

Correspondence and requests for materials should be addressed to A.A. (e-mail: adriano@pathol.unizh.ch).

Remission in models of type 1 diabetes by gene therapy using a single-chain insulin analogue

Hyun Chul Lee^{*†}, Su-Jin Kim^{*†}, Kyung-Sup Kim^{†‡}, Hang-Cheol Shin^{*} & Ji-Won Yoon^{*§}

^{*} Division of Endocrinology, Department of Internal Medicine; [†] Institute of Genetic Science, Department of Biochemistry and Molecular Biology, Yonsei University, College of Medicine, Seoul 120–752, Korea
[§] Laboratory of Viral and Immunopathogenesis of Diabetes, Julia McFarlane Diabetes Research Centre, Department of Microbiology and Infectious Diseases, Faculty of Medicine, The University of Calgary, Calgary, Alberta T2N 4N1, Canada

[†] These authors contributed equally to the work.

A cure for diabetes has long been sought using several different approaches, including islet transplantation, regeneration of β cells and insulin gene therapy¹. However, permanent remission of type 1 diabetes has not yet been satisfactorily achieved. The development of type 1 diabetes results from the almost total destruction of insulin-producing pancreatic β cells by autoimmune responses specific to β cells^{2–6}. Standard insulin therapy may not maintain blood glucose concentrations within the relatively narrow range that occurs in the presence of normal pancreatic β cells⁷. We used a recombinant adeno-associated virus (rAAV) that expresses a single-chain insulin analogue (SIA), which possesses biologically active insulin activity without enzymatic conversion, under the control of hepatocyte-specific L-type pyruvate kinase (LPK) promoter, which regulates SIA expression in response to blood glucose levels. Here we show that SIA produced from the gene construct rAAV-LPK-SIA caused remission of diabetes in streptozotocin-induced diabetic rats and autoimmune diabetic mice for a prolonged time without any apparent side effects. This new SIA gene therapy may have potential therapeutic value for the cure of autoimmune diabetes in humans.

First, we generated SIA by replacing 35 residues of the C-peptide with a short turn-forming heptapeptide (Gly-Gly-Gly-Pro-Gly-Lys-Arg). We produced recombinant SIA in *Escherichia coli*, folded it and examined its biological activity. We found that the *in vitro* receptor binding and glucose uptake activity and the *in vivo* hypoglycaemic effect of SIA were substantially greater than that of proinsulin, but slightly lower than insulin, indicating that the biological activity of SIA is comparable to that of insulin (Table 1).

Second, we constructed a recombinant plasmid, pLPK-SIA

Table 1 Functional properties of the recombinant single-chain insulin analogue

	Insulin receptor binding* (ED ₅₀ in nM)	Glucose uptake† (ED ₅₀ in nM)	<i>In vivo</i> hypoglycaemic effect‡ (ED ₅₀ in nmol kg ^{–1})	
			1 hour	2 hours
Human insulin	0.742 (100%)§	0.322 (100%)	1.5 ± 0.34 (100%)	1.8 ± 0.37 (100%)
Human proinsulin	36.1 (2%)	26.5 (1.25%)	9.1 ± 1.12 (16.5%)	9.6 ± 1.28 (18.8%)
Single-chain insulin analogue (SIA)	2.68 (27.7%)	1.56 (21.3%)	3.6 ± 0.42 (41.6%)	3.4 ± 0.18 (52.9%)

* ED₅₀ values are presented as the concentration of unlabelled insulin or analogues required to decrease tracer insulin binding to 50% of maximal.

† ED₅₀ values are presented as the concentration of insulin or analogues resulting in half-maximal glucose uptake rate.

‡ ED₅₀ values are presented as the dose of insulin or analogues that gave half of the maximal hypoglycaemic activity 1 or 2 h after subcutaneous administration to fasted rats.

§ The percentage of proinsulin and SIA were determined on the basis of the ED₅₀ value of human insulin as 100%.

(Fig. 1a) by cloning the SIA gene under the LPK promoter^{8–12}. The albumin leader sequence was added to the frame of the SIA gene construct to facilitate the secretion of SIA from the hepatocytes, and the simian virus 40 (SV40) enhancer was added downstream of the SIA gene in order to elevate the basal level of SIA expression. When we administered pLPK-SIA into the portal vein of streptozotocin (STZ)-induced diabetic SD rats, we found that the blood glucose levels gradually decreased until the fifth day after administration, remained in a normoglycaemic state for an additional four days and gradually rose thereafter to reach over 500 mg dl⁻¹ at 14 days after treatment. In contrast, STZ-induced diabetic rats treated with the pSIA construct without LPK remained hyperglycaemic (Fig. 1b). When we determined the presence of SIA DNA and mRNA in the liver, kidney, spleen, lung and heart five days after plasmid administration, we found that the transfected plasmid DNA was detected in only liver cells from both groups of rats, whereas SIA mRNA was expressed in only the pLPK-SIA-treated rats (Fig. 1c).

Third, we attempted to overcome the short half-life of the transfected DNA using an adeno-associated virus (AAV) vector, which is a safe and efficient gene delivery system^{13,14} that integrates into the host chromosomal DNA in a site-specific manner^{15–17}. We produced recombinant AAV containing the LPK promoter-SIA DNA (rAAV-LPK-SIA) in 293 cells. To determine whether rAAV-LPK-SIA can remit diabetes, we administered 10¹¹ particles per animal through the portal vein of STZ-induced diabetic SD rats. The blood glucose levels gradually decreased in the rAAV-LPK-SIA-treated rats, reached the level of normoglycaemia one week after treatment and remained in a normoglycaemic state for more than eight months (Fig. 2a). However, lower doses of rAAV-LPK-SIA (10⁹–10¹⁰ virus particles) were insufficient for complete remission. Higher doses of rAAV-LPK-SIA (10¹² virus particles per rat) did not result in hypoglycaemia in the treated rats (Fig. 2a).

We then examined the molecular state of the rAAV genome in the

liver of the recipients by Southern blots of liver DNA with SIA complementary DNA and the SV40 sequence as a probe. We found that rAAV-LPK-SIA DNA existed in both single-stranded and double-stranded states at the early stage (5 days) after the treatment, but only in the double-stranded state at later stages (5–25 weeks). In addition, we found that rAAV-LPK-SIA-DNA was integrated into the hepatocyte chromosomal DNA in a head-to-tail concatemeric manner at 25 weeks after treatment (Fig. 2b). These results indicate that the single-stranded recombinant adeno-associated viral genome is converted to a double-stranded genome in the infected hepatocytes and integrated into the chromosomal DNA.

We examined the localization of SIA expression in the liver by anti-SIA antibody staining, and found that SIA was expressed in the hepatocytes of rats treated with rAAV-LPK-SIA (Fig. 3a), but not in untreated control rats (Fig. 3b). SIA-expressing cells were clustered around the central vein and portal triads in the liver (Fig. 3a). Histological examination of the infected liver sections revealed no detectable damage (Fig. 3c and d). In addition, we found no difference in the level of plasma aspartate transaminase (AST) or plasma alanine transaminase (ALT), marker enzymes for hepatic damage, between saline-treated, normal control rats (94 ± 12 IU l⁻¹ for AST and 46 ± 9 IU l⁻¹ for ALT; *n* = 5) and rAAV-LPK-SIA-treated rats (95 ± 10 IU l⁻¹ for AST and 40 ± 9 IU l⁻¹ for ALT). Furthermore, we did not find anti-SIA antibody in the sera of the treated rats during the experimental period (data not shown). These results indicate that treatment of diabetic rats with rAAV-LPK-SIA resulted in the remission of diabetes without any apparent change in the infected liver cells or any adverse effects of SIA expression on the hepatocytes.

We questioned whether the remission of diabetes might be due to the secretion of insulin from residual pancreatic β cells in the STZ-

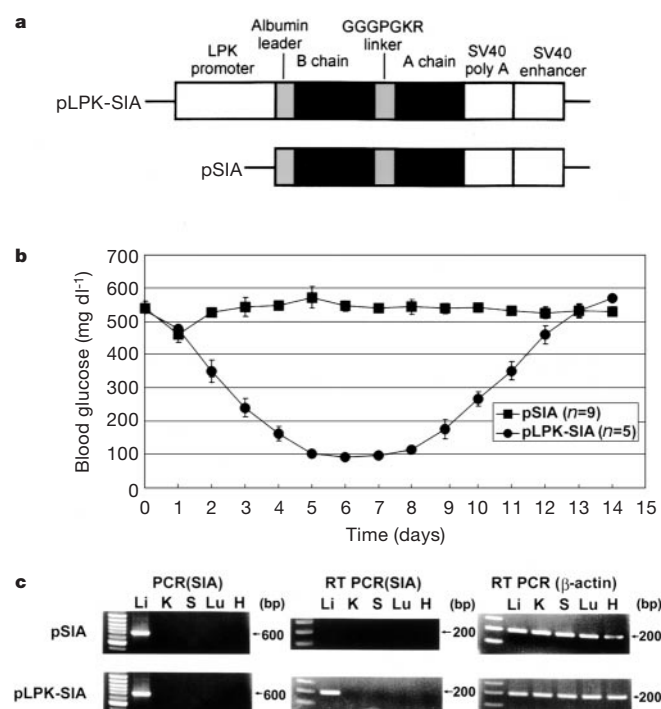


Figure 1 Construction of pLPK-SIA, hypoglycaemic effect of pLPK-SIA and hepatocyte-specific expression of pLPK-SIA. **a**, Diagram of pLPK-SIA and pSIA. **b**, Blood glucose levels in diabetic rats injected with pSIA or pLPK-SIA (mean ± standard deviation, s.d.). **c**, SIA DNA (left), SIA RNA (centre) or β-actin RNA as an internal control (right) in the liver (Li), kidney (K), spleen (S), lung (Lu) and heart (H) of rats 5 days after treatment with pSIA or pLPK-SIA. bp, base pairs.

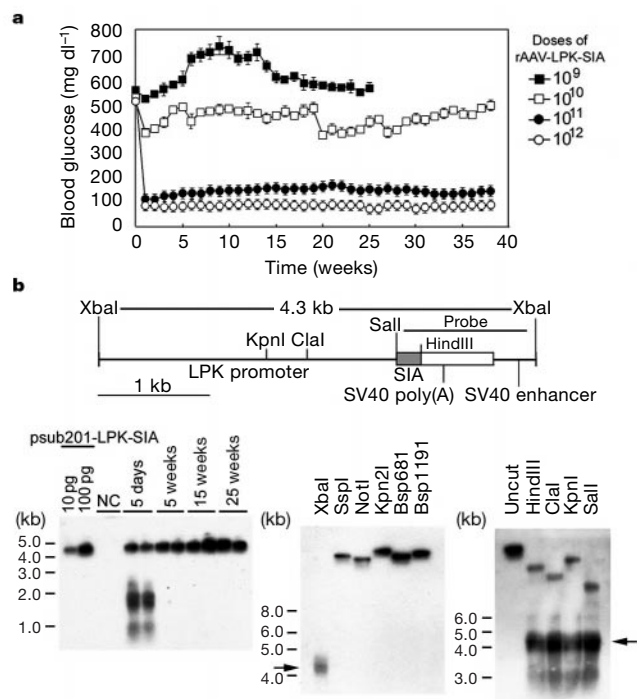


Figure 2 Hypoglycaemic effect of rAAV-LPK-SIA and integration of rAAV-LPK-SIA DNA into hepatocyte DNA. **a**, Blood glucose levels in STZ-induced diabetic SD rats (*n* = 10 per group) after injection of different doses of rAAV-LPK-SIA (mean ± s.d. value). **b**, Restriction enzyme map of LPK-SIA and the region used as a probe for Southern blots (top). Southern blot hybridization of liver DNA from rats treated with rAAV-LPK-SIA after digestion with XbaI at different times after treatment (left) and digestion with non-cutting (centre) or single cutting (right) enzymes at 25 weeks after treatment. XbaI-digested psub 201-LPK-SIA and liver DNA from normal SD rats (NC) were used as controls; arrows, 4.3-kb band.

treated rats, rather than the expression of SIA. We stained islets of rAAV-LPK-SIA-treated rats with anti-insulin antibody, and found that insulin-producing β cells were rarely seen in these islets (Fig. 3e), whereas insulin-producing β cells were clearly detected in normal control islets (Fig. 3f). In contrast, SIA was continuously produced for over 8 months after the treatment with rAAV-LPK-SIA (Fig. 3g; data after 25 weeks not shown). In addition, a negligible amount of C-peptide was found in plasma of the rAAV-LPK-SIA-treated rats after glucose loading (Fig. 3h). These results indicate that the control of blood glucose in rAAV-LPK-SIA-treated rats was not due to endogenous pancreatic insulin, but to the expression of SIA in the liver cells.

Fourth, we questioned whether the expression of SIA is truly regulated by blood glucose levels through the LPK promoter. We maintained different levels of blood glucose (approximately 100, 300 or 500 mg dl⁻¹) by a hyperglycaemic clamp method¹⁸ for 30 min in rats in which the blood glucose had been normalized after treatment with 10¹¹ particles of rAAV-LPK-SIA, and examined the production of SIA in the hepatocytes and plasma four hours after glucose loading. We found that the level of SIA expression was closely correlated with the concentration of blood glucose (Fig. 4b and c), even though similar levels of SIA DNA were found among

the different groups of rats (Fig. 4a), indicating that the expression of SIA under the LPK promoter is properly controlled by blood glucose levels.

We next determined whether rAAV-LPK-SIA-treated rats (15–17 weeks old) clear glucose from the blood in a manner similar to non-diabetic normal rats by glucose tolerance tests (GTTs; Fig. 4d). We injected glucose (2 g per kg body weight intraperitoneally (i.p.), after 4 h of fasting) into rats treated with rAAV-LPK-SIA that had recovered from diabetes or normal, untreated rats and examined the plasma insulin or SIA levels and blood glucose concentrations after glucose loading. We found that the plasma insulin levels of normal rats peaked within 30 min and returned to basal levels at 2 h after glucose loading (Fig. 4e). However, the plasma SIA levels of rAAV-LPK-SIA-treated rats were delayed, peaking at 3–4 h and returning to basal levels at 6 h after glucose loading (Fig. 4e and f). The expression of SIA mRNA in the hepatocytes showed a similar pattern to that of plasma SIA (Fig. 4g). This difference in kinetics between insulin and SIA in response to glucose loading is probably due to the fact that insulin is rapidly released into the blood by exocytosis from the pancreatic β cells, which are highly sensitive to small changes in extracellular glucose. In contrast, the secretion of SIA is controlled at the transcriptional level by the LPK promoter;

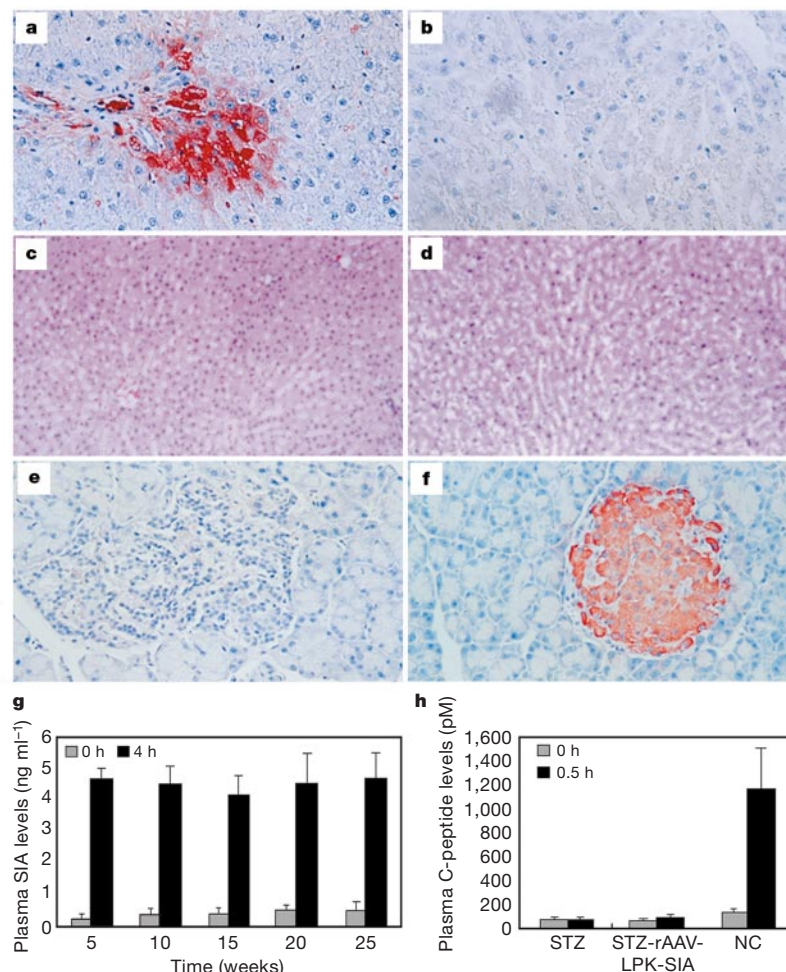


Figure 3 Expression of SIA in hepatocytes and plasma of rAAV-LPK-SIA-treated rats. Anti-SIA antibody staining of liver sections from **a**, rAAV-LPK-SIA-treated rats at 1 month after treatment and **b**, normal control rats. Haematoxylin and eosin (HE) staining of liver tissue from **c**, rAAV-LPK-SIA-treated rats and **d**, normal control rats. Anti-insulin antibody staining of pancreatic islets from **e**, rAAV-LPK-SIA-treated rats and **f**, normal control rats.

g, Plasma SIA levels at 0 and 4 h after glucose loading in rAAV-LPK-SIA-treated rats ($n = 7$ per group). **h**, Plasma C-peptide levels at 0 and 0.5 h after glucose loading at 25 weeks after rAAV-LPK-SIA treatment (STZ-rAAV-LPK-SIA; $n = 7$). Normal nondiabetic rats (NC; $n = 7$) and untreated STZ-induced diabetic rats (STZ; $n = 7$); mean $\pm$ s.d.

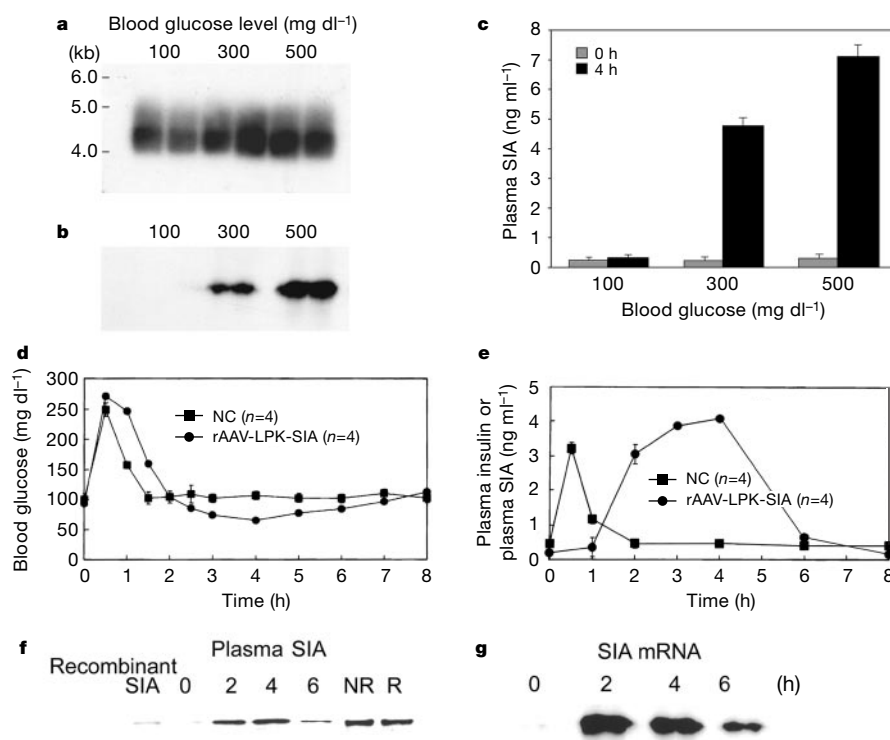


Figure 4 Response of SIA to glucose in rAAV-LPK-SIA-treated STZ-induced diabetic rats. **a**, Presence of SIA DNA and **b**, expression of SIA in the hepatocytes, and **c**, production of SIA in the plasma ($n = 7$ per group) after the maintenance of different blood glucose levels at 30 weeks after treatment with rAAV-LPK-SIA. **d**, Glucose and **e**, insulin or SIA levels during GTTs performed 4 weeks after treatment. NC indicates normal control rats;

rAAV-LPK-SIA indicates rats treated with rAAV-LPK-SIA (mean $\pm$ s.d.). **f**, Western blots of plasma SIA at the indicated times (h) after glucose loading. NR, non-reducing condition; R, reducing condition. **g**, Expression of SIA mRNA after glucose loading examined by RNase protection assay.

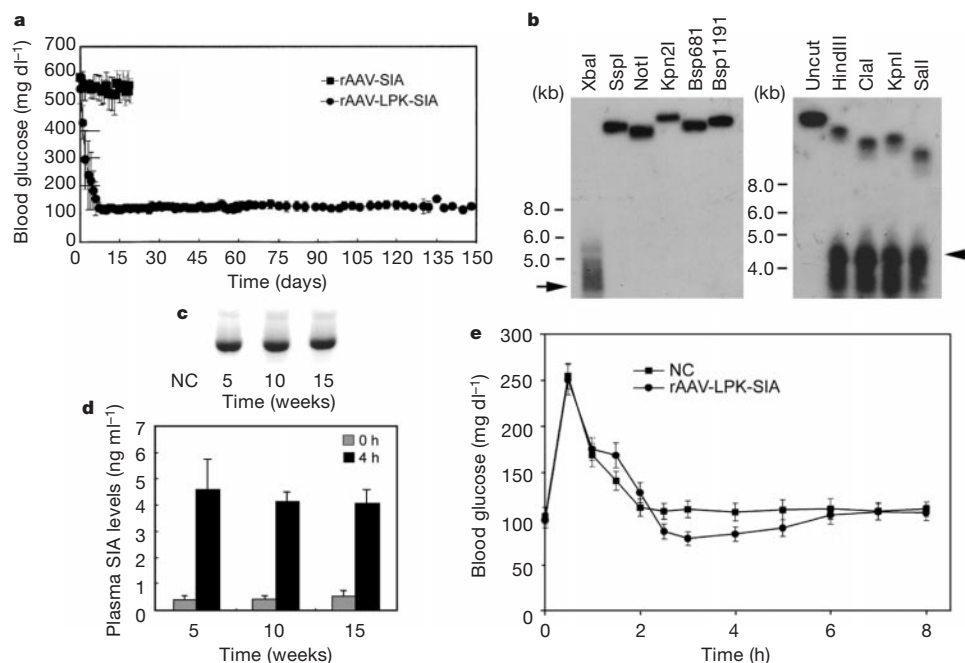


Figure 5 Remission of autoimmune diabetes in NOD mice by administration of rAAV-LPK-SIA. **a**, Blood glucose levels in diabetic NOD mice after rAAV-SIA or rAAV-LPK-SIA treatment ($n = 6$ per group). **b**, Integration of SIA DNA into hepatocyte DNA analysed by Southern blot at 15 weeks after rAAV-LPK-SIA treatment (arrows, 4.3-kb band). **c**, SIA

mRNA in hepatocytes (NC, untreated NOD mice) and **d**, plasma SIA at 0 and 4 h after glucose loading at various times after treatment ($n = 5$ per group). **e**, Blood glucose levels during GTTs in rAAV-LPK-SIA-treated NOD mice ($n = 5$) and control NOR mice (NC, $n = 7$); mean $\pm$ s.d.

thus a prolonged period is required to change the plasma SIA level in response to blood glucose.

In these GTTs, we found that the blood glucose levels of normal rats peaked at 30 min after glucose loading, returned to the normal range (around 100 mg dl⁻¹) at 90 min, and stabilized thereafter (Fig. 4d). The blood glucose levels of rats treated with rAAV-LPK-SIA showed a similar pattern to normal rats, except for a slightly delayed time to recover normal glucose levels (120 min versus 90 min) and transiently lower blood glucose levels from 3 to 6 h after glucose loading (Fig. 4d). Although the time required for plasma SIA levels to peak was delayed by 3.5 h, as compared to insulin, the recovery to normoglycaemia was only delayed by 30 min, as compared to the normal controls. This result indicates that the action of SIA on blood glucose levels began when SIA concentrations were rising during the first hour after glucose loading; thus the rising levels of SIA and the falling levels of blood glucose to normoglycaemia are well correlated within the first 2 h. The subsequent elevation of SIA from 2 to 5 h did not result in significant hypoglycaemia. We speculate that this may be due to a reduction in glucose sensitivity to SIA when the blood glucose levels are normalized, a reduction of the sensitivity of the post-receptor signal cascade during the period of high SIA levels, and/or compensatory increases in glucagon and catecholamines, which can increase blood glucose.

Fifth, we attempted to find whether rAAV-LPK-SIA can also cause remission of autoimmune diabetes in non-obese diabetic (NOD) mice. When we administered rAAV-LPK-SIA (10¹² virus particles per mouse) to diabetic NOD mice, we found that the blood glucose levels reached normoglycaemia at seven days after treatment and remained in a normoglycaemic state for more than five months. In contrast, diabetic NOD mice treated with rAAV-SIA (without the LPK promoter) remained hyperglycaemic and died within three weeks (Fig. 5a). When we examined the presence of the SIA gene in the hepatocytes of NOD mice at 15 weeks after treatment with rAAV-LPK-SIA, we found that SIA DNA was integrated into the chromosomal DNA (Fig. 5b). We then examined SIA expression in the liver and plasma of the mice after glucose loading at 5, 10, and 15 weeks after treatment with rAAV-LPK-SIA, and found that SIA mRNA was clearly expressed (Fig. 5c) and SIA protein was released in the plasma (Fig. 5d). Anti-SIA antibody was not detected in the sera from these mice during the five months after treatment (data not shown). In addition, we performed GTTs in NOD mice that had recovered from diabetes. We found that the blood glucose levels of NOD mice treated with rAAV-LPK-SIA peaked at 30 min after glucose loading, returned to the normal range (115 mg dl⁻¹) at 120 min and stabilized thereafter (Fig. 5e). The time to recover normal blood glucose levels was slightly delayed and the blood glucose levels from 3 to 6 hours after glucose loading were slightly lower compared with normal non-obese diabetes-resistant (NOR) mice, similar to the pattern seen in the rats treated with rAAV-LPK-SIA.

We have developed a potential method for the treatment of autoimmune type 1 diabetes by the expression of a single-chain insulin analogue in the hepatocytes under the control of a hepatocyte-specific glucose regulatable promoter and SV40 enhancer. The host's cell-mediated autoimmune responses do not attack the SIA-expressing hepatocytes, resulting in long-term remission of autoimmune diabetes in NOD mice. The treatment of both chemically induced diabetes in rats and autoimmune diabetes in NOD mice with rAAV expressing the insulin analogue resulted in the permanent remission of type 1 diabetes without any detectable adverse effect on the hepatocytes. □

Methods

Cloning and expression of single-chain insulin analogue cDNA in *E. coli*

SIA cDNA was generated by polymerase chain reaction (PCR) using five overlapping oligonucleotides of 65–68 bases in length. The resulting SIA cDNA sequence was:

ATG/TTC/GTT/AAT/CAG/CAC/CTG/TGC/GGC/TCT/CAC/CTG/GTA/GAA/GCT/CTG/TAC/CTG/GTT/TGC/GGT/GAA/CGT/GGT/TTT/TTC/TAC/ACC/CCG/AAA/ACC/GGT/GGT/GGT/CCG/CGT/AAA/CGT/GGC/ATC/GTT/GAA/CAA/TGC/TGT/ACT/AGC/ATC/TGC/TCT/CTC/TAC/CAG/CTG/GAG/AAC/TAT/TGT/AAAC/TAG/TAA. The amino-terminal pentapeptide sequence (PSDKP) of TNF- α was added as a fusion partner. Ten histidine residues for convenience in the purification process and a methionine residue for chemical cleavage were inserted between the PSDKP sequence and SIA, then cloned into pET-3a¹⁹ under the T7 promoter (pET-SIA). The fused SIA protein was expressed in *E. coli* BL21 (DE3) as inclusion bodies. The fusion protein was sulphonated at the cysteine residues and cleaved by CNBr treatment. S-sulphonated SIA was purified by cation-exchange chromatography (Pharmacia Biotech), refolded with β -mercaptoethanol and analysed by reverse-phase HPLC.

Examination of functional activity of recombinant SIA produced in *E. coli*

Insulin receptor binding and glucose uptake assays were performed using IM-9 lymphocytes as described previously^{20–22}. To determine the *in vivo* hypoglycaemic activity of SIA, 8–10-week-old male Sprague–Dawley rats were fasted, and SIA protein (4–80 μ g in 0.1 ml saline per 100 g body weight) or the same volume of saline as a control was injected subcutaneously²³. Blood was obtained from the tail vein of each rat, and the glucose level was determined.

Construction of pSIA and pLPK-SIA

The SIA cDNA from pET-SIA was subcloned into the PCR-script (Stratagene) at the *Bam*HI/*Hind*III site. The SV40 poly(A) signal sequence from pCDM8 (Invitrogen) and the SV40 enhancer from the pGL3 (Promega) were amplified by PCR and subcloned at the *Hind*III/*Apa*I and *Apa*I sites, respectively. The albumin leader sequence (72 base pairs) was inserted in front of the SIA cDNA using a ExSite PCR-based Site-Directed Mutagenesis Kit (Stratagene). The clone containing the albumin leader sequence was isolated and designated as pSIA. The final construct, pLPK-SIA, was generated by insertion of the promoter of the rat LPK gene (–3193 to +18) amplified by PCR into pSIA at the *Xba*I/*Sall* site.

Administration of pSIA, pLPK-SIA, rAAV-SIA or rAAV-LPK-SIA into the liver

Animals were anaesthetized with Ketamine-chloride (10 mg kg⁻¹). A midline abdominal incision was made, and a mixture of 30 μ g of DNA-FuGENE 6 (Boehringer Mannheim) and rAAV-SIA or rAAV-LPK-SIA (10⁹–10¹² virus particles) was injected into the hepatic portal vein of animals that had been hyperglycaemic for 2 weeks.

PCR and PCR after reverse transcription of RNA

PCR was performed using the sense primers derived from the T7 or LPK promoter region (5'-GTAATACGACTCACTATAG GGC-3' for pSIA-injected rats; 5'-ATTTCGAATAAG AAGAGGAAGGGAAG-3' for rats injected with pLPK-SIA) and the antisense primers derived from the 3' terminus of the SIA gene (5'-GCGCAAGCTTTTACTAGTTACAAT AGTT-3'). To detect SIA mRNA, the total RNA was isolated from various tissues and RT-PCR was performed using the primers 5'-GCGCGGATCCATGTTCTGTTAATCAGCAC-3' and 5'-GCGCAAGCTTTTACTAGTTACAATAGTT-3'. β -actin mRNA was amplified as an internal control.

Production of recombinant AAV-SIA and AAV-LPK-SIA

The 4.3-kilobase LPK-SIA including the SV40 enhancer was amplified by PCR from the pLPK-SIA plasmid and subcloned into psub201 at the *Xba*I site (psub201-LPK-SIA). The recombinant AAV (rAAV-LPK-SIA) was produced and purified as described previously^{13–15}. The psub 201-SIA (without LPK) was used for the construction of rAAV-SIA. The amount of viral DNA was quantified by competitive PCR¹⁴.

Southern blot analysis

Liver DNA (10 μ g) was digested with various restriction enzymes. After agarose gel electrophoresis, the DNA was transferred to a membrane, hybridized with a ³²P-labelled probe containing the SIA cDNA and SV40 sequence, and detected by autoradiography.

Measurement of plasma insulin, C-peptide and SIA and blood glucose

Plasma insulin and C-peptide levels were measured using the Linco rat insulin RIA kit and Linco rat C-peptide RIA kit, respectively (Linco Research Immunoassay). For measurement of plasma SIA levels, we performed competitive ELISA using rabbit polyclonal anti-SIA antibodies. Blood glucose levels were measured as described elsewhere²⁴.

Hyperglycaemic clamp experiment

A 20% glucose solution, for maintenance of blood glucose at about 300 or 500 mg dl⁻¹, or saline, for maintenance of blood glucose at about 100 mg dl⁻¹, was perfused into the femoral vein using an electronic digital syringe pump (Harvard)¹⁸. The infusion rate was adjusted to maintain the desired blood glucose levels for 30 min. The production of SIA in the hepatocytes and plasma was examined 4 h later.

Western blot analysis

Proteins below 30,000 daltons were separated and concentrated from 1.5 ml of pooled plasma at 0, 2, 4 and 6 h. The proteins, with or without β -mercaptoethanol treatment, were electrophoretically separated on a 10–20% Tricine gradient gel (NOVEX). Western blot analysis was performed using rabbit polyclonal anti-SIA antibody.

RNase protection assay

10 µg of liver RNA was used for RNase protection analysis with ³²P-labelled SIA antisense RNA made from pET-SIA by *in vitro* transcription using T7 RNA polymerase.

Measurement of AST and ALT

AST and ALT were measured in an autochemical analyser (Hitachi 747) using an ultraviolet assay method²⁵.

Histological and immunohistochemical analysis

Sections of liver and pancreas were histologically examined^{26,27} and stained with anti-SIA or anti-insulin antibodies, respectively^{27,28}.

Received 7 August; accepted 9 October 2000.

- Levine, F. & Leibowitz, G. Towards gene therapy of diabetes mellitus. *Mol. Med. Today* **5**, 165–171 (1999).
- Yoon, J. W. & Jun, H. S. In *Encyclopedia of Immunology* (eds Roitt, I. M. & Delves, P. J.) 2nd edn, 1390–1398 (Academic, London, 1998).
- Schranz, D. B. & Lernmark, A. Immunology in diabetes: an update. *Diab. Metab. Rev.* **14**, 3–29 (1998).
- Tisch, R. & McDevitt, H. Insulin-dependent diabetes mellitus. *Cell* **85**, 291–297 (1996).
- Bach, J. F. Insulin-dependent diabetes mellitus as a β cell targeted disease of immunoregulation. *J. Autoimmun.* **8**, 439–463 (1995).
- Rossini, A. A., Greiner, D. L., Friedman, H. P. & Mordes, J. P. Immunopathogenesis of diabetes mellitus. *Diabetes Rev.* **1**, 43–75 (1993).
- The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New Engl. J. Med.* **329**, 977–986 (1993).
- Cuif, M. H., Doiron, B. & Kahn, A. Insulin and cyclic AMP act at different levels on transcription of the L-type pyruvate kinase gene. *FEBS Lett.* **417**, 81–84 (1997).
- Chen, R., Doiron, B. & Kahn, A. Glucose responsiveness of a reporter gene transduced into hepatocytic cells using a retroviral vector. *FEBS Lett.* **365**, 223–226 (1995).
- Decaux, J. F., Antoine, B. & Kahn, A. Regulation of the expression of the L-type pyruvate kinase gene in adult rat hepatocytes in primary culture. *J. Biol. Chem.* **264**, 11584–11590 (1989).
- Cuif, M. H., Porteu, A., Kahn, A. & Vaulont, S. Exploration of a liver-specific, glucose/insulin-responsive promoter in transgenic mice. *J. Biol. Chem.* **268**, 13769–13772 (1993).
- Bergot, M. O., Diaz-Guerra, M. J., Puzeat, N., Raymondjean, M. & Kahn, A. Cis-regulation of the L-type pyruvate kinase gene promoter by glucose, insulin and cyclic AMP. *Nucleic Acids Res.* **20**, 1871–1877 (1992).
- Muzyczka, N. Use of adeno-associated virus as a general transduction vector for mammalian cells. *Curr. Top. Microbiol. Immunol.* **158**, 97–129 (1992).
- Clark, K. R., Liu, X., McGrath, J. P. & Johnson, P. R. Highly purified recombinant adeno-associated virus vectors are biologically active and free of detectable helper and wild-type viruses. *Hum. Gene Ther.* **10**, 1031–1039 (1999).
- Samulski, R. J. Adeno-associated virus: integration at a specific chromosomal locus. *Curr. Opin. Genet. Dev.* **3**, 74–80 (1993).
- Giraud, C., Winocour, E. & Berns, K. I. Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence. *Proc. Natl Acad. Sci. USA* **91**, 10039–10043 (1994).
- Kotin, R. M., Linden, R. M. & Berns, K. I. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J.* **11**, 5071–5078 (1992).
- Cameron, N. E., Cotter, M. A. & Low, P. A. Nerve blood flow in early experimental diabetes in rats: relation to conduction deficits. *Am. J. Physiol.* **261**, E1–E8 (1991).
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89 (1990).
- Pollet, R. J., Standaert, M. L. & Haase, B. A. Insulin binding to the human lymphocyte receptor. Evaluation of the negative cooperativity model. *J. Biol. Chem.* **252**, 5828–5834 (1977).
- Roth, J. Assay of peptide hormones using cell receptors: application to insulin and to human growth hormone. *Methods Enzymol.* **37**, 66–82 (1975).
- Frost, S. C. & Lane, M. D. Evidence for the involvement of vicinal sulfhydryl groups in insulin-activated hexose transport by 3T3-L1 adipocytes. *J. Biol. Chem.* **260**, 2646–2652 (1985).
- Heath, W. F. et al. (A-C-B) human proinsulin, a novel insulin agonist and intermediate in the synthesis of biosynthetic human insulin. *J. Biol. Chem.* **267**, 419–425 (1992).
- Yoon, J. W., Lesniak, M. A., Fussganger, R. & Notkins, A. L. Genetic differences in susceptibility of pancreatic β -cells to virus-induced diabetes mellitus. *Nature* **264**, 178–180 (1976).
- Ma, Z. et al. Effect of hemoglobin- and Perflubron-based oxygen carriers on common clinical laboratory tests. *Clin. Chem.* **43**, 1732–1737 (1997).
- Yoon, J. W., Rodrigues, M. M., Currier, C. & Notkins, A. Long-term complications of virus-induced diabetes mellitus in mice. *Nature* **296**, 566–569 (1982).
- Yoon, J. W. et al. Control of autoimmune diabetes in NOD mice by GAD expression or suppression in cells. *Science* **284**, 1183–1187 (1999).
- Hirasawa, K. Possible role of macrophage-derived soluble mediators in the pathogenesis of EMC virus-induced diabetes in mice. *J. Virol.* **71**, 4024–4031 (1997).

Acknowledgements

We thank R. J. Samulski for providing psub201 and pXX2, A. L. Kyle and K. Clarke for editorial assistance and B. Pinder for artwork. We also thank H. S. Jun for a critical review of the manuscript. J. W. Y. is a Heritage Medical Scientist awardee of the Alberta Heritage Foundation for Medical Research. This work was supported in part by Brain Korea 21 Project.

Correspondence and requests for materials should be addressed to H.C.L. (e-mail: endohcllee@yumc.yonsei.ac.kr)

A ubiquitin-like system mediates protein lipidation

Yoshinobu Ichimura^{*†‡}, Takayoshi Kirisako^{*†‡}, Toshifumi Takao[§], Yoshinori Satomi[§], Yasutsugu Shimonishi[§], Naotada Ishihara^{*}, Noboru Mizushima^{*||}, Isei Tanida[¶], Eiki Kominami[¶], Mariko Ohsumi[#], Takeshi Noda^{*†} & Yoshinori Ohsumi^{*†}

^{*} Department of Cell Biology, National Institute for Basic Biology, 38 Nishigonaka, Myodaijicho, Okazaki 444-8585, Japan

[†] Department of Molecular Biomechanics, School of Life Science, The Graduate University for Advanced Studies, Okazaki 444-8585, Japan
[‡] Institute for Protein Research, Osaka University, Yamadaoka 3-2, Suita, Osaka 565-0871, Japan

^{||} PRESTO, Japan Science and Technology Corporation (JST), Kawaguchi 332-0012, Japan

[#] Department of Biosciences, High Tech Research Center, Teikyo University of Science and Technology, Yamanashi 409-0193, Japan

[¶] Department of Biochemistry, Juntendo University School of Medicine, Tokyo 113-8421, Japan

[‡] These authors contributed equally to this work

Autophagy is a dynamic membrane phenomenon for bulk protein degradation in the lysosome/vacuole^{1,2}. Apg8/Aut7 is an essential factor for autophagy in yeast^{3–5}. We previously found that the carboxy-terminal arginine of nascent Apg8 is removed by Apg4/Aut2 protease, leaving a glycine residue at the C terminus⁶. Apg8 is then converted to a form (Apg8-X) that is tightly bound to the membrane⁶. Here we report a new mode of protein lipidation. Apg8 is covalently conjugated to phosphatidylethanolamine through an amide bond between the C-terminal glycine and the amino group of phosphatidylethanolamine. This lipidation is mediated by a ubiquitination-like system. Apg8 is a ubiquitin-like protein that is activated by an E1 protein, Apg7 (refs 7, 8), and is transferred subsequently to the E2 enzymes Apg3/Aut1 (ref. 9). Apg7 activates two different ubiquitin-like proteins, Apg12 (ref. 10) and Apg8, and assigns them to specific E2 enzymes, Apg10 (ref. 11) and Apg3, respectively. These reactions are necessary for the formation of Apg8-phosphatidylethanolamine. This lipidation has an essential role in membrane dynamics during autophagy⁶.

Apg8 was the first molecule found to localize to the intermediate structures of the autophagosome, and is necessary for autophagosome formation^{4,5}. Apg8-X behaves like an integral membrane protein, even though Apg8 does not possess a membrane-spanning region⁶. Here we used phase partitioning with Triton X-114 to separate Apg8-X. In wild-type cells, a small amount of Apg8 was detected in the detergent phase—most of it was in the aqueous phase (Fig. 1a). In contrast, *Δapg7* cells resulted in the exclusive partitioning of Apg8 to the aqueous phase (Fig. 1a). Apg8-X accumulated in *Δapg4Δapg8* cells expressing Apg8FG (Apg8 lacking the C-terminal arginine)⁶, resulting in marked enhancement of detergent phase partitioning (Fig. 1a). These results indicate that Apg8 acquires sufficient hydrophobicity for membrane insertion after the attachment of molecule X.

To assess directly the structure of Apg8-X, we purified it from *Δapg4Δapg8* cells expressing His₆-Apg8FG. Matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOFMS) gave two signals at *m/z* 14,368 and 15,003 (Fig. 1b). The former value was assigned to the molecule His₆-Apg8FG (14,297) modified with an acrylamide monomer (relative molecular mass (*M_r*) 71) derived from SDS/polyacrylamide gel electrophoresis (PAGE). The latter was assigned to the His₆-Apg8-X. The difference in molecular mass (635) and hydrophobic nature of Apg8-X suggested that X is a glycerophospholipid. In fact, the His₆-Apg8-

X signal shifted to a lower molecular size on treatment with mild alkali, a process that removes fatty acids from glycerophospholipids (data not shown).

We elucidated the attachment site and exact chemical structure of X by obtaining a C-terminal fragment of His₆-Apg8-X after mild alkaline treatment and subsequent digestion by lysyl endopeptidase, and subjected it to electrospray ionization/orthogonal tandem mass spectrometry. Every resulting fragment containing the C-terminal Gly was heavier by M_r 197 than the corresponding fragment from Apg8FG (compare two series of y'' ions in Fig. 1c). This value (197) is consistent with the mass of a glycerophosphoethanolamine moiety. In addition, some y'' ions from His₆-Apg8-X accompanied signals reduced by M_r 172, which can be explained by loss of a glycerophosphate moiety. We also found using gas chromatography/mass spectrometry that saturated (C16:0 and C18:0) and unsaturated (C16:1 and C18:1) fatty acids were liberated from His₆-Apg8-X by mild alkaline treatment (data not shown). From the above evidence, we concluded that phosphatidylethanolamine (PE) is covalently conjugated to the C-terminal glycine of Apg8. To our knowledge, this is the first protein lipidation in which a protein conjugates to a phosphoglycerolipid commonly found in biological membranes.

Next, we explored the mechanism of this lipidation of Apg8. We previously reported that this conversion of Apg8 to Apg8-PE requires Apg7, an E1 enzyme in the Apg12 conjugation system^{7,8,10,11}. Apg8 and Apg12 show weak homology in their C-terminal regions (Fig. 2a). This suggested that Apg8 may be a ubiquitin-like protein (Ubl) and that Apg7 might function as an E1 enzyme for Apg8. First, we assessed whether Apg7 and Apg8 interact directly. Apg8 was co-immunoprecipitated with Apg7-Myc using anti-Myc antibody (Fig. 2b). Conversely, Apg7-Myc was also co-immunoprecipitated with haemagglutinin (HA)-Apg8 using anti-HA antibody (data not shown). This interaction disappeared in Δ apg4 cells, and recovered with the expression of Apg8FG, but not Apg8F, which lacks the C-terminal glycine (Fig. 2c). These results show the dependence of the Apg7-Apg8 interaction upon the C-terminal glycine of Apg8.

Apg7 requires Cys 507 and Gly 333 for the activation of Apg12 (ref. 7). We found that the same residues of Apg7 participate in its interaction with Apg8, because the interaction was reduced by the substitution of those residues to serine or alanine (C507S, C507A and G333A) (data not shown). Furthermore, we detected HA-Apg8-Apg7-Myc complexes with both anti-HA and anti-Myc

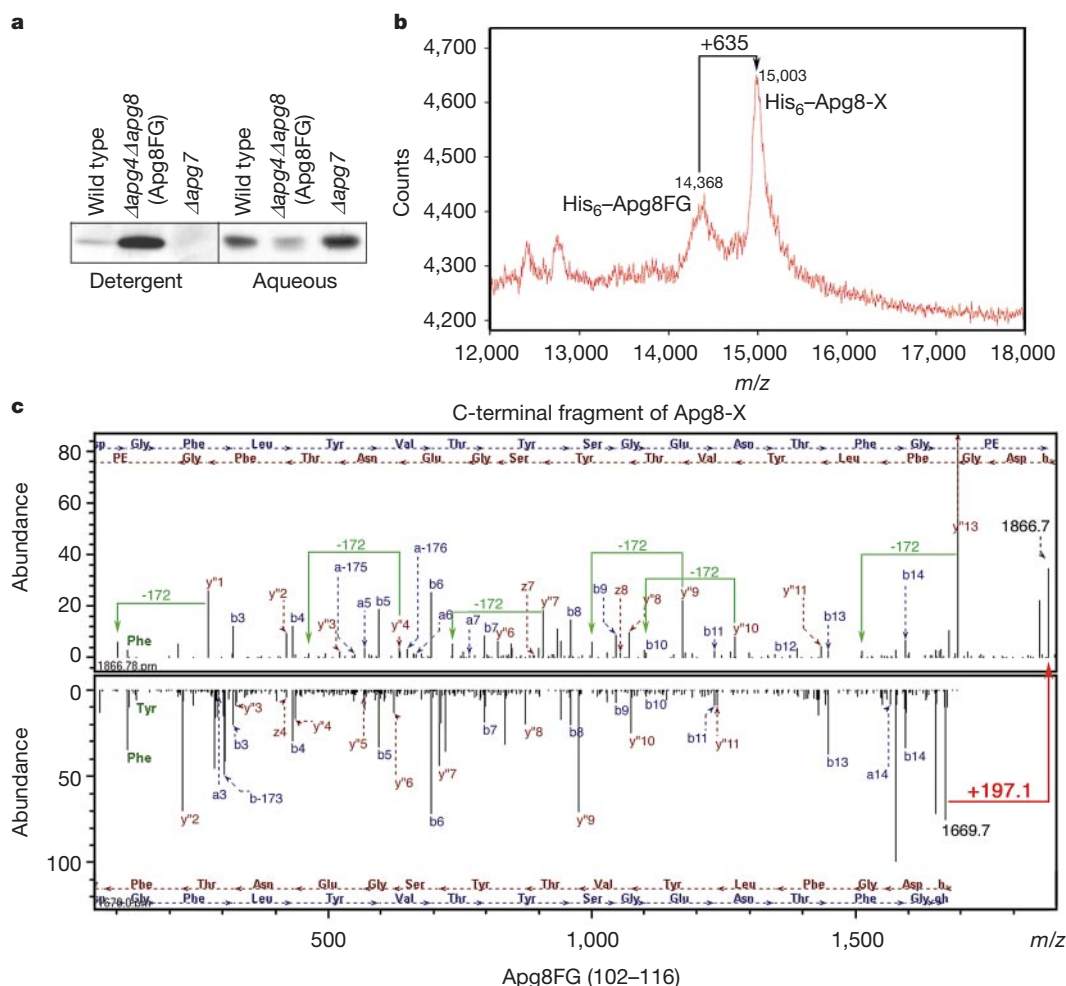


Figure 1 Characterization of molecule X. **a**, Phase partitioning of Apg8. Wild-type and Δ apg4 Δ apg8 cells producing Apg8FG and Δ apg7 cells were subjected to phase partitioning (see Methods). **b**, MALDI mass spectrum of His₆-Apg8. Purified His₆-Apg8 (predominantly Apg8FG-X and a small amount of Apg8FG) was extracted with 0.4% β -octylglycoside from a polyacrylamide gel and desalted with ZipTipC₄ (Millipore) before mass measurement. The principal MH⁺ ion signal at m/z 15,003 is shifted in mass by about +635 from the MH⁺ value corresponding to His₆-Apg8FG. **c**, MS/MS spectra of the C-terminal peptides of Apg8FG-X (top) and Apg8FG (bottom). C-terminal fragments were

generated from Apg8FG-X and Apg8FG (see Methods). Each fragment that was observed only as non-¹⁸O-labelled signals was subjected to MS/MS. The MS/MS spectra, deconvoluted with MaxEnt3, were interpreted by SeqMS software²⁶. Arrows show the sequences from the N and C terminus based on y''_m and b_l ions, respectively (where m and l denote positions counted from the C and N terminus) that were produced by cleavage of peptide bonds during MS/MS. Amino acids shown in three-letter code denote immonium ions. The nomenclature of these ions is as in previous work²⁸.

antibodies in the immunoprecipitates of HA-Apg8 after non-reducing SDS-PAGE, but not in the presence of dithiothreitol (DTT) (Fig. 2d). Apg8 appears to be linked to Apg7 through a thioester bond between the C-terminal glycine of Apg8 and Cys 507 of Apg7. These results indicate that Apg8 is a Ubl that is activated by the E1 enzyme Apg7. The three-dimensional structure of GATE-16, a mammalian homologue of Apg8, has recently been reported. In support of our results, that structure exhibited an overall similarity to the ubiquitin fold¹².

We next searched for an E2 enzyme for Apg8. Among autophagy-defective (*apg*) mutants¹³, we isolated *APG3*. Sequence analysis revealed that *APG3* is identical to *AUT1*, which is essential for autophagy but has an unknown function⁹. The region around Cys 234 of Apg3 is partially homologous to the region surrounding the active centre (Cys 133) of Apg10, an E2 enzyme in the Apg12 system¹¹, as well as those of mammalian homologues (Fig. 3a). Because of the weak homology, we assumed that Apg3/Aut1 was a candidate for E2 and examined its physical interaction with Apg8 (Fig. 3a). Myc-Apg8 was pulled down successfully by anti-Apg3 antibody (Fig. 3b). We constructed two mutant forms of Apg3, Apg3C234S and Apg3C234A. Neither mutant was detectable in the Myc-Apg8 immunoprecipitates with anti-Myc antibody (Fig. 3c). Furthermore, Myc-Apg8-Apg3 conjugates were detected in the Myc-Apg8 immunoprecipitates using both anti-Myc and anti-Apg3 antibodies in the non-reducing conditions, but not in the presence of DTT (Fig. 3d). These results indicate that Apg8 conjugates with Apg3 through a thioester bond between the C-terminal glycine of Apg8 and Cys 234 of Apg3. The Apg8-Apg3 conjugate was not found in Δ *apg7* cells, but observable in Δ *apg10* cells (Fig. 3d). Apg3 should be a specific E2 enzyme for Apg8.

In Δ *apg4*, Δ *apg7* and Δ *apg3* cells, Apg8-PE was not detected in SDS-PAGE containing 6 M urea, in which it migrates faster than Apg8 (Fig. 4a). Thus, the Apg8 ubiquitination-like system must be requisite for Apg8-PE formation. Next, we examined the relevance of Apg8-PE on autophagy. In the *apg3*^{C234A} mutant, the progression of autophagy was found to be defective (Fig. 4b). It was also

defective in the Cvt pathway, a biosynthetic membrane transport similar to autophagy (data not shown)². We previously showed that the C-terminal glycine of Apg8, and Cys 507 and Gly 333 of Apg7, are essential for the process^{6,7}. Formation of Apg8-PE, then, must be required for autophagy and the Cvt pathway.

We have identified a new ubiquitin-like system. Most intriguingly, Apg8 is conjugated to a lipid, phosphatidylethanolamine

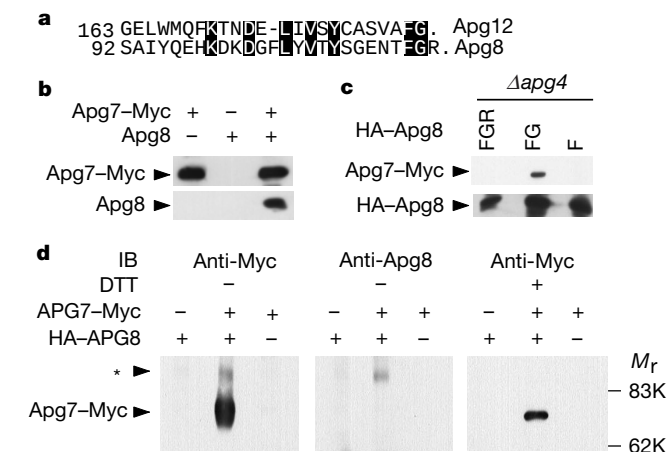


Figure 2 Apg8 is activated by Apg7 (E1 in the Apg12 Ubl system). **a**, Homology between Apg8 and Apg12 at the C termini. **b**, Interaction between Apg8 and Apg7. Δ *apg7* Δ *apg8* cells were transformed with APG7-Myc and/or APG8 on 2 μ plasmids. Cell lysates were immunoprecipitated with anti-Myc antibody, and immunoblotted with anti-Myc or anti-Apg8 antibody. Apg8 co-immunoprecipitated with Apg7-Myc. **c**, Effect of Apg8 C-terminal processing on the Apg8-Apg7 interaction. Δ *apg4* cells co-expressing Apg7-Myc and HA-Apg8 (FGR, FG and F) from 2 μ plasmids. FGR, wild-type Apg8; FG, Apg8 lacking C-terminal arginine; F, Apg8 lacking C-terminal glycine and arginine. Cell lysates were immunoprecipitated with anti-HA antibody, and immunoblotted with anti-HA and anti-Myc antibody. **d**, Thioester formation between Apg8 and Apg7. Cell extracts of Δ *apg7* Δ *apg8* cells expressing Apg7-Myc and/or HA-Apg8 from 2 μ plasmids were immunoprecipitated with anti-HA antibody, and immunoblotted (IB) with anti-Myc and anti-HA antibody. DTT-, no DTT; DTT+, 100 mM DTT; asterisk, Apg8-Apg7 conjugate.

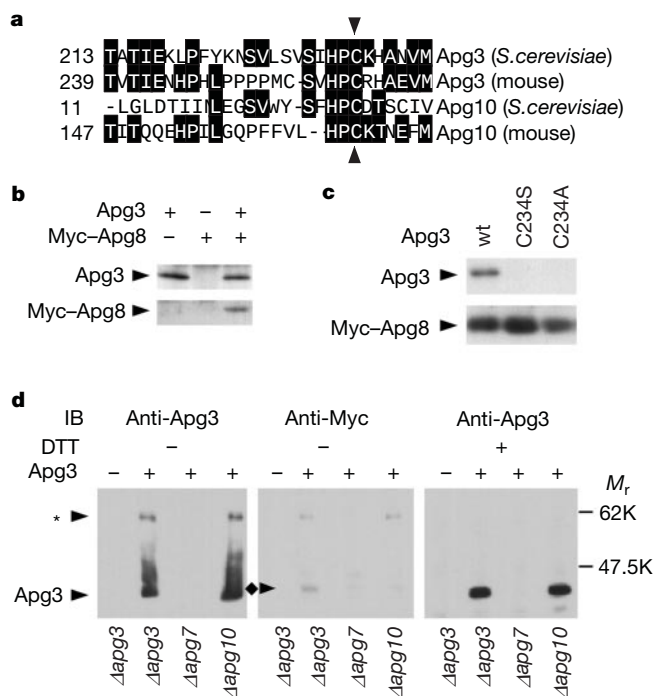


Figure 3 Apg3 is a conjugating enzyme (E2) for Apg8. **a**, Homology between Apg3, Apg10 and their homologues. Cys 234 of Apg3 corresponds to the active centre Cys 133 of Apg10 (arrowhead). **b**, Interaction between Apg8 and Apg3. Lysates of Δ *apg3* cells expressing APG3 and/or Myc-APG8 from 2 μ plasmids were immunoprecipitated with anti-Apg3 antibody, and immunoblotted with anti-Apg3 and anti-Myc antibody. **c**, Effect of mutated Cys 234 in Apg3 on the interaction between Apg8 and Apg3. Lysates of Δ *apg3* cells expressing Myc-APG8 and either wild-type (wt) APG3 or mutant APG3 (C234S and C234A) from 2 μ plasmids were immunoprecipitated with anti-Myc antibody, and immunoblotted with anti-Myc and anti-Apg3 antibodies. **d**, Thioester formation between Apg8 and Apg3. Δ *apg3*, Δ *apg7* and Δ *apg10* cells were transformed with both Myc-APG8 and Apg3, or only Apg3, on 2 μ plasmids. Cell extracts were immunoprecipitated with anti-Myc antibody, and subjected to immunoblot (IB) using anti-Myc and anti-Apg3 antibody. DTT-, no DTT; DTT+, 100 mM DTT; asterisk, Apg8-Apg3 conjugate; diamond, possible Myc-Apg8 dimer.

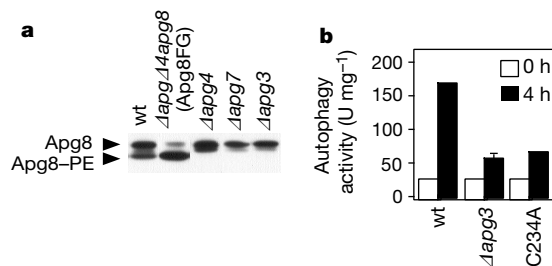


Figure 4 Apg8 ubiquitination-like system is essential for autophagy. **a**, Apg7 and Apg3 are necessary for attachment of Apg8 to PE. Δ *apg* cell lysates were subjected to SDS-PAGE containing 6 M urea, and immunoblotted with anti-Apg8 antibody. **b**, Autophagic activity in *apg3* mutants. Wild-type, Δ *apg3* and *apg3*^{C234A} cells were grown in rich medium (open bar) and subjected to nitrogen starvation for 4 h (filled bar). The autophagic activity of each cell was assessed by measuring the alkaline phosphatase activity. The *apg3*^{C234A} mutant showed defective autophagy.

(PE), not to a protein, although the details of the lipid conjugation event after Apg8–Apg3 conjugation remain to be resolved. A few observations indicate that an amide bond links Apg8 to PE. First, mass spectroscopy indicates that the glycine residue of Apg8 is conjugated to an ethanolamine moiety. Second, the primary amine of PE resembles the side chain of lysine, the usual target site for ubiquitin and Ubl system substrates^{14–16}. Last, like ubiquitin and Ubl systems, PE is cleaved from the C-terminal glycine of Apg8 by the processing enzyme, Apg4 (ref. 6). Ubiquitin is modified with phospholipid in baculo viral membrane, but it may not be relevant to our findings because the lipid is not attached to the C-terminal glycine¹⁷. The Apg8 system constitutes a new type of protein lipidation. This reversible lipidation is necessary for membrane dynamics in autophagy. In higher eukaryotes, Apg8 comprises a protein family including GATE-16, which stimulates the ATPase activity of NSF (N-ethylmaleimide-sensitive fusion protein) in intra-Golgi transport and LC3 in autophagy^{18–21}. It is possible that the homologues are also Ubls attached to lipids such as PE. This type of lipidation might be required not only for autophagy but also for various membrane dynamics. □

Methods

Yeast strains

We used the *Saccharomyces cerevisiae* parental strains YW5-1B (MATa *ura3 leu2 trp1*), SEY6210 (MATα *ura3 leu2 trp1 his3 lys2 suc2*) and their derivatives. The APG3 gene was cloned by complementation of an *apg3* mutant, MT38-4-3 (MATa *ura3 leu2 apg3-1*)¹³ as described²².

Genetic methods

We carried out gene manipulation and yeast genetics according to described methods^{23,24}. Site-directed mutagenesis of APG3 and APG7 was performed by using QuikChange (Stratagene).

Immunoblotting

Whole-cell lysates were prepared by disruption with glass beads in a lysis buffer as described⁴. Protein was quantified by BCA (Pierce). SDS–PAGE and immunoblot analyses were performed as described⁴.

Phase partitioning

We converted cells to spheroplasts as described⁴. Spheroplasts were lysed with PBS containing 2% Triton X-114. After extraction overnight on ice, phase partitioning was performed with multiple washings²⁵.

Isolation of Apg8-X

A 13,000g pellet of *Δapg4Δapg8* cells expressing His₆–Apg8FG was prepared after full growth⁶. After washing with 1 M NaCl, the pellet was treated with 1% Triton X-100, followed by centrifugation at 100,000g for 30 min. The resulting supernatant was affinity purified on a Ni-NTA column followed by an anti-Apg8 antibodies column. Polyacrylamide gel fragments containing His₆–Apg8-X (and a small amount of His₆–Apg8FG) were prepared by SDS–PAGE.

Preparation of C-terminal fragments of Apg8FG and Apg8FG-X

Polyacrylamide gels containing recombinant Apg8FG produced by *Escherichia coli* and containing purified His₆–Apg8 were dried and treated with 0.06 M NaOH in 30% CH₃OH at 40 °C for 4 h to hydrolyse acyl bonds between glycerol and fatty acids. Proteins were digested in gel with lysyl endopeptidase (1 pmol) in 50 mM Tris–HCl (pH 8.0) prepared with 40 atom per cent H₂¹⁸O at 37 °C for 12 h. The resulting peptides were extracted with 0.1% trifluoroacetic acid (TFA) in 50% acetonitrile from the gel and desalted with ZipTipC₁₈ (Millipore) before mass measurement.

Mass spectrometry (MS)

Matrix-assisted laser-desorption/ionization MS was performed using a Voyager Elite XL time-of-flight mass spectrometer. A protein sample was mixed with the matrix solution, the supernatant of a 50% acetonitrile solution saturated with α-cyano-4-hydroxycinnamic acid, and then air dried on the flat surface of a stainless steel plate. Ions were generated by irradiating the sample area with the output of a nitrogen laser at a wavelength of 337 nm. Electrospray ionization (ESI)–MS/MS was carried out on a hybrid quadrupole orthogonal acceleration tandem mass spectrometer (Q-TOF) (Micromass, Manchester, UK). Solutions containing peptides were loaded into a borosilicate nanoflow tip (Micromass), and set into an ESI source. MS/MS data were processed by the maximum entropy data enhancement program MaxEnt 3 (Micromass), which is capable of deconvoluting a spectrum in which peaks of varying charge states are present, thus producing a simplified spectrum comprising only mono-isotopic peaks in a single charge state. The resultant spectra were interpreted by SeqMS²⁶, a software aid for *de novo* sequencing by MS/MS.

Immunoprecipitation

Cells were suspended in immunoprecipitation buffer (50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA, 1 mM PMSF, 1% NP-40 and protease inhibitor cocktail (Roche Molecular Biochemicals)) and disrupted by using glass beads. The lysates were incubated with anti-Myc (9E10), anti-HA (16B12) (BAbCo), or anti-Apg3 antibody at 4 °C for 2 h, after which 20 μl of Protein G/Sepharose or Protein A/Sepharose (50% slurry) (Amersham Pharmacia Biotech) was added. After incubation at 4 °C for 2 h, the immunoprecipitates were collected by centrifugation. The immunocomplexes bound to the Sepharose beads were washed three times with immunoprecipitation buffer, and eluted from the beads with SDS–PAGE sample buffer (50 mM Tris–HCl (pH 6.8), 2% SDS, 10% glycerol and 0.2% bromophenol blue, with or without 100 mM DTT) at 65 °C for 5 min. Samples were separated by SDS–PAGE, and analysed by immunoblotting.

Alkaline phosphatase (ALP) assay

The APG3 gene was disrupted in TN125 (MATa *ura3 leu2 trp1 his3 lys2 ade2 PHO8::pho8Δ60*), and the ALP assay was done as described²⁷.

Antibodies

Full-length Apg3 was expressed in *E. coli* and purified by HiTrap Chelating (Amersham Pharmacia), and then used for immunization of rabbits according to conventional procedures. Anti-Myc (9E10) and anti-HA (16B12) antibodies were purchased from BAbCo. Production of anti-Apg8 antibody has been described⁴. Anti-API antibody was a gift from D. J. Klionsky (Univ. Michigan, Ann Arbor).

Received 24 July; accepted 11 October 2000.

- Baba, M., Takeshige, K., Baba, N. & Ohsumi, Y. Ultrastructural analysis of the autophagic process in yeast: detection of autophagosomes and their characterization. *J. Cell Biol.* **124**, 903–913 (1994).
- Klionsky, D. J. & Ohsumi, Y. Vacuolar import of proteins and organelles from the cytoplasm. *Annu. Rev. Cell Dev. Biol.* **15**, 1–32 (1999).
- Lang, T. et al. Aut2p and Aut7p, two novel microtubule-associated proteins are essential for delivery of autophagic vesicles to the vacuole. *EMBO J.* **17**, 3597–3607 (1998).
- Kirisako, T. et al. Formation process of autophagosome is traced with Apg8/Aut7p in yeast. *J. Cell Biol.* **147**, 435–446 (1999).
- Huang, W. P., Scott, S. V., Kim, J. & Klionsky, D. J. The itinerary of a vesicle component, Aut7p/Cvt5p, terminates in the yeast vacuole via the autophagy/Cvt pathways. *J. Biol. Chem.* **275**, 5845–5851 (2000).
- Kirisako, T. et al. The reversible modification regulates the membrane-binding state of Apg8/Aut7 essential for autophagy and the cytoplasm to vacuole targeting pathway. *J. Cell Biol.* **151**, 263–276 (2000).
- Tanida, I. et al. Apg7p/Cvt2p: A novel protein-activating enzyme essential for autophagy. *Mol. Biol. Cell* **10**, 1367–1379 (1999).
- Kim, J., Dalton, V. M., Eggerton, K. P., Scott, S. V. & Klionsky, D. J. Apg7p/Cvt2p is required for the cytoplasm-to-vacuole targeting, macroautophagy, and peroxisome degradation pathways. *Mol. Biol. Cell* **10**, 1337–1351 (1999).
- Schlumpberger, M. et al. AUT1, a gene essential for autophagocytosis in the yeast *Saccharomyces cerevisiae*. *J. Bacteriol.* **179**, 1068–1076 (1997).
- Mizushima, N. et al. A protein conjugation system essential for autophagy. *Nature* **395**, 395–398 (1998).
- Shintani, T. et al. Apg10p, a novel protein-conjugating enzyme essential for autophagy in yeast. *EMBO J.* **18**, 5234–5241 (1999).
- Paz, Y., Elazar, Z. & Fass, D. Structure of GATE-16, membrane transport modulator and mammalian ortholog of autophagocytosis factor Aut7p. *J. Biol. Chem.* **275**, 25445–25450 (2000).
- Tsukada, M. & Ohsumi, Y. Isolation and characterization of autophagy-defective mutants of *Saccharomyces cerevisiae*. *FEBS Lett.* **333**, 169–174 (1993).
- Hochstrasser, M. Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405–439 (1996).
- Ciechanover, A. The ubiquitin-proteasome pathway: on protein death and cell life. *EMBO J.* **17**, 7151–7160 (1998).
- Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
- Guarino, L. A., Smith, G. & Dong, W. Ubiquitin is attached to membranes of baculovirus particles by a novel type of phospholipid anchor. *Cell* **80**, 301–309 (1995).
- Wang, H., Bedford, F. K., Brandon, N. J., Moss, S. J. & Olsen, R. W. GABA_A-receptor-associated protein links GABA_A receptors and the cytoskeleton. *Nature* **397**, 69–72 (1999).
- Legesse-Miller, A., Sagiv, Y., Glzman, R. & Elazar, Z. Aut7p, a soluble autophagic factor, participates in multiple membrane trafficking processes. *J. Biol. Chem.* **275**, 32966–32973 (2000).
- Sagiv, Y., Legesse-Miller, A., Porat, A. & Elazar, Z. GATE-16, a membrane transport modulator, interacts with NSF and the Golgi v-SNARE GOS-28. *EMBO J.* **19**, 1494–1504 (2000).
- Kabeya, Y. et al. LC3, a mammalian homolog of yeast Apg8p, is localized in autophagosome membranes after processing. *EMBO J.* **19**, 5720–5728 (2000).
- Kametaka, S., Matsuura, A., Wada, Y. & Ohsumi, Y. Structural and functional analyses of APG5, a gene involved in autophagy in yeast. *Gene* **178**, 139–143 (1996).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).
- Rose, M. D., Winston, F. & Hieter, P. *Methods in Yeast Genetics: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1990).
- Brusca, J. S. & Radolf, J. D. Isolation of integral membrane proteins by phase partitioning with Triton X-114. *Methods Enzymol.* **228**, 182–193 (1994).
- Fernandez-de-Cossio, J. et al. Automated interpretation of high-energy collision-induced dissociation spectra of singly protonated peptides by 'SeqMS', a software aid for *de novo* sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **12**, 1867–1878 (1998).
- Noda, T. & Ohsumi, Y. Tor, a phosphatidylinositol kinase homologue, controls autophagy in yeast. *J. Biol. Chem.* **273**, 3963–3966 (1998).
- Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T. & Watson, J. T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **59**, 2621–2625 (1987).

Acknowledgements

We would like to thank D. J. Klionsky for the anti-API antibody, A. Kihara for technical advice, and Jun-ichi Osuga for analysis of fatty acids by GC-MS. This work was supported in part by Grants-in-Aid for the Ministry of Education, Science, Culture and Sports of Japan.

Correspondence and requests for materials should be addressed to Y.O.
(e-mail: yohsumi@nibb.ac.jp).

$G\alpha_i$ and $G\alpha_o$ are target proteins of reactive oxygen species

Motohiro Nishida*, Yoshiko Maruyama*, Rie Tanaka*, Kenji Kontani†, Taku Nagao* & Hitoshi Kurose*

* Laboratory of Pharmacology and Toxicology, † Laboratory of Physiological Chemistry, Graduate School of Pharmaceutical Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Reactive oxygen species (ROS) have been identified as central mediators in certain signalling events^{1–4}. In the heart, ROS have important functions in ischaemia/reperfusion-induced cardiac injury^{5,6} and in cytokine-stimulated hypertrophy⁷. Extracellular signal-regulated kinase (ERK) is one of the ROS-responsive serine/threonine kinases. Previous studies showed that tyrosine kinases and small G proteins are involved in the activation of ERK by ROS^{4,8}; however, the initial target protein of ROS that leads to ERK activation remains unknown. Here we show that inhibition of the $\beta\gamma$ -subunit of G protein ($G\beta\gamma$) attenuates hydrogen peroxide (H_2O_2)-induced ERK activation in rat neonatal cardio-

myocytes. The $G\beta\gamma$ -responsive ERK activation induced by H_2O_2 is independent of ligands binding to G_i -coupled receptors, but requires phosphatidylinositol-3-kinase and Src activation. In *in vitro* studies, however, treatment with H_2O_2 increases [³⁵S]GTP- γ S binding to cardiac membranes and directly activates purified heterotrimeric G_i and G_o but not G_s . Analysis using heterotrimeric G_o and its individual subunits indicates that H_2O_2 modifies $G\alpha_o$ but not $G\beta\gamma$, which leads to subunit dissociation. We conclude that $G\alpha_i$ and $G\alpha_o$ are critical targets of oxidative stress for activation of ERK.

Extracellular signal-related kinase (ERK) is a member of the mitogen-activated protein (MAP) kinase family that is activated by H_2O_2 , an oxidative stress. Several reports have indicated similarities between signalling components involved in H_2O_2 -induced activation and those involved in G_i -coupled receptor-mediated activation. For instance, Src and Ras have been activated by both H_2O_2 treatment and the stimulation of G_i -coupled receptors^{4,8,9}.

We found that H_2O_2 activated ERK in rat neonatal cardio-

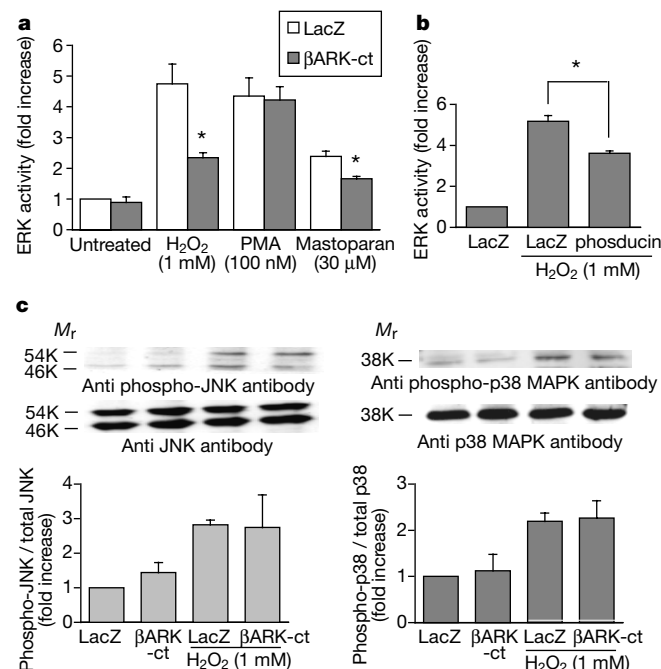


Figure 1 Characteristics of H_2O_2 -induced activation of ERK. **a**, **b**, Effect of LacZ (control), β ARK-ct (**a**) or phosducin (**b**) on ERK activities of H_2O_2 -, PMA- and mastoparan-treated myocytes. ERK activities are shown as fold increase over unstimulated activity. Treatments marked with an asterisk are significantly different from control (LacZ) ($P < 0.05$). **c**, Effect of β ARK-ct on H_2O_2 -induced activation of JNK and p38 MAPK. Neonatal myocytes were exposed to 1 mM H_2O_2 for 10 min, and JNK and p38 MAPK activation was determined.

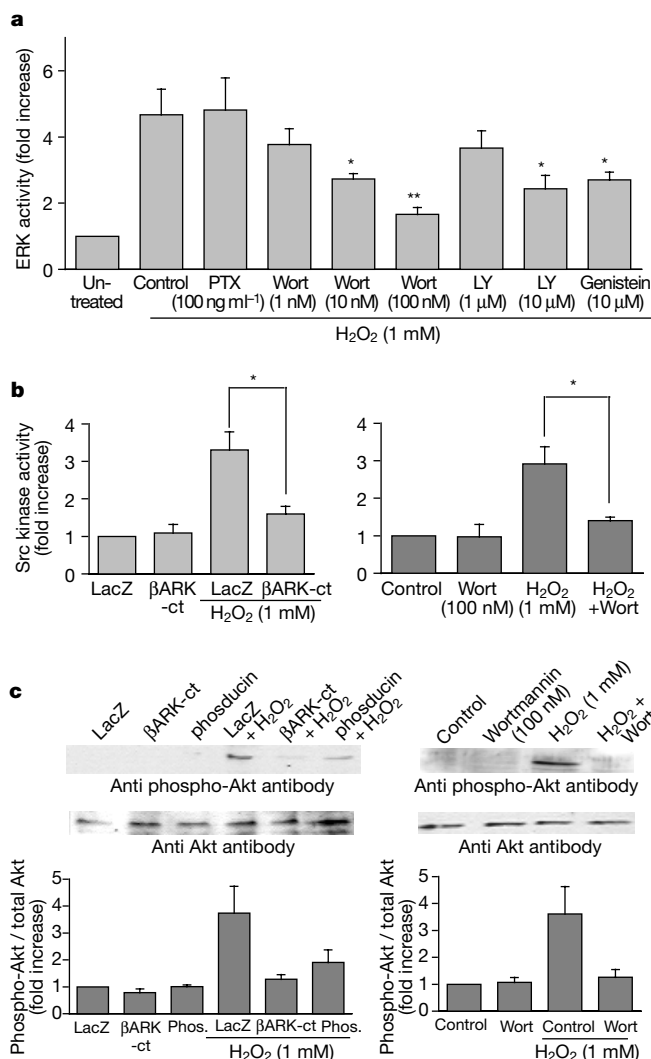


Figure 2 Effect of β ARK-ct on activation of signalling components that participate in H_2O_2 -induced activation of ERK. **a**, Neonatal myocytes were treated with pertussis toxin (PTX), wortmannin (wort), LY294002 (LY) or genistein, and the H_2O_2 -induced ERK activity was determined. Wortmannin, LY294002 and genistein were added 20 min before H_2O_2 treatment, and PTX was added 18 h before H_2O_2 treatment. **b**, Src activity in cells expressing β ARK-ct (left) or cells treated with 100 nM wortmannin (right). **c**, Effect of β ARK-ct, phosducin and 100 nM wortmannin on Akt activation. Significance from control (LacZ): asterisk, $P < 0.05$; double asterisk, $P < 0.01$.

myocytes in a dose-dependent manner (see Supplementary Information). To determine the involvement of G protein in this process, we expressed the carboxy terminus of β -adrenergic receptor kinase (β ARK-ct) by the adenoviral gene-transfer method^{10–12} to inhibit the function of G $\beta\gamma$. Expression of β ARK-ct inhibits the activation of several G $\beta\gamma$ -regulated effectors by sequestering G $\beta\gamma$ (ref. 13). The level of β ARK-ct in neonatal cardiomyocytes by recombinant adenovirus was about 40 times higher than in the endogenous expression level of G $\beta\gamma$ ($8.43 \pm 1.84 \text{ ng } \mu\text{g}^{-1} \text{ protein}$, β ARK-ct versus $0.22 \pm 0.01 \text{ ng } \mu\text{g}^{-1} \text{ protein}$, G $\beta\gamma$). The expression of β ARK-ct significantly attenuated H₂O₂-induced activation of ERK but did not affect activation by phorbol 12-myristate 13-acetate (PMA), a direct activator of protein kinase C (Fig. 1a). The direct activation of heterotrimeric G protein by mastoparan (30 μM) activated ERK, and the expression of β ARK-ct attenuated it.

Our β ARK-ct construct contains the pleckstrin homology domain, as well as the G $\beta\gamma$ -binding domain. As the pleckstrin homology domain can bind phosphatidylinositol-4,5-bisphosphates (PIP₂), the inhibitory effect of β ARK-ct on H₂O₂-induced activation of ERK may be caused by binding to PIP₂ rather than to G $\beta\gamma$. To exclude this possibility, we expressed another G $\beta\gamma$ inhibitory protein, phosducin¹⁴. The expression of phosducin in neonatal cardiomyocytes also attenuated H₂O₂-induced ERK activation (Fig. 1b). The difference in the magnitude of the inhibition of ERK activation might be explained by the different affinities of these proteins for G $\beta\gamma$ ¹⁵, by the lower expression of phosducin as compared with β ARK-ct or by the incomplete sequestration of G $\beta\gamma$ by β ARK-ct. However, our results suggest that β ARK-ct has a

specific action on ERK activation, and that β ARK-ct attenuates H₂O₂-induced activation of ERK by inhibiting the function of G $\beta\gamma$ in heterotrimeric G proteins.

H₂O₂ also activated the other stress-responsive MAP kinases (MAPKs), c-Jun amino-terminal kinase/ stress activated protein kinase (JNK/SAPK) and p38 MAPK 2–3 fold. However, this activation was not significantly affected by the expression of β ARK-ct (Fig. 1c), suggesting that in the primary culture of rat neonatal cardiomyocytes, JNK/SAPK and p38 MAPK are activated by H₂O₂ by means of a G $\beta\gamma$ -independent process.

Treatment with 100 ng ml⁻¹ of pertussis toxin (PTX) did not attenuate H₂O₂-induced ERK, indicating that the effect of H₂O₂ is not mediated by a released factor acting on G_i-coupled receptors (Fig. 2a). Treatment with wortmannin or LY294002, inhibitors of phosphatidylinositol 3-kinase (PI(3)K), significantly attenuated H₂O₂-induced ERK activation in a dose-dependent manner. Treatment with genistein, a tyrosine kinase inhibitor, also attenuated ERK activation induced by H₂O₂. These results indicate that both PI(3)K and tyrosine kinase are involved in the H₂O₂-induced activation of ERK.

We next examined whether the site of action of β ARK-ct is upstream of Src—a genistein-sensitive tyrosine kinases. Treatment of H₂O₂ increased Src kinase activity by threefold (Fig. 2b). Activation of Src was significantly inhibited by the expression of β ARK-ct or by treatment with wortmannin, indicating that G $\beta\gamma$ and PI(3)K function upstream of Src in H₂O₂-induced ERK activation.

We then examined the relationship between G $\beta\gamma$ and PI(3)K in H₂O₂-induced ERK activation. Oxidative stress activates PI(3)K, leading to Akt phosphorylation, and Akt is an important mediator of ERK activation by various stimuli in different cells¹⁶. We found that treatment with H₂O₂ increased the phosphorylated form of Akt by threefold (Fig. 2c). This increase was attenuated by the expression of β ARK-ct or phosducin. This result indicates that the site of action of these proteins is upstream of Akt. As expected, H₂O₂-induced activation of Akt was also inhibited by the PI(3)K

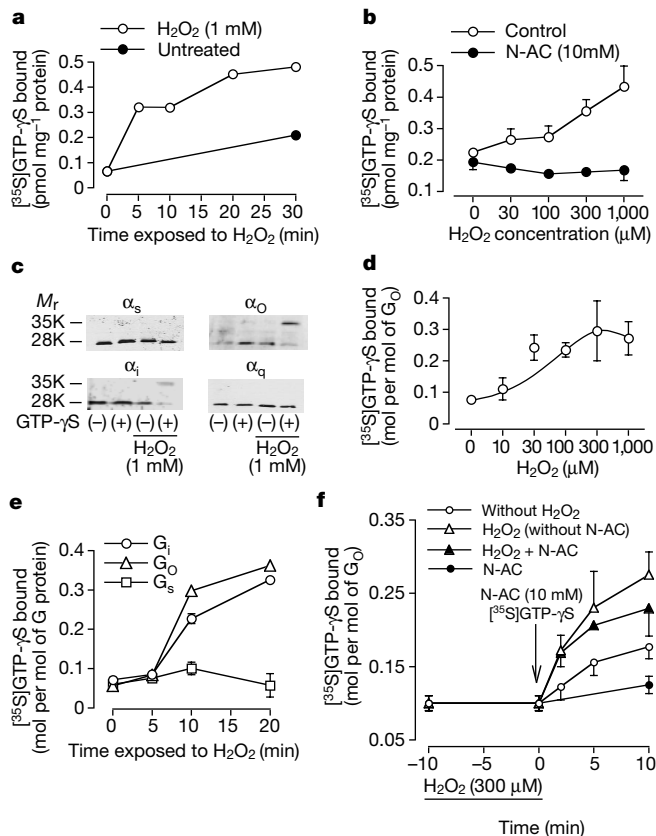


Figure 3 GTP- γ S binding to crude membrane and purified G proteins. **a**, **b**, [³⁵S]GTP- γ S binding to the membrane preparation was determined for the indicated time in the absence (**a**) or presence (**b**) of 10 mM *N*-acetyl-L-cysteine (NAC). **c**, The CHAPS-solubilized membranes were treated with H₂O₂ and 50 μM GTP- γ S, and subjected to trypsin sensitivity assay. **d**, [³⁵S]GTP- γ S binding to G₀ at differing concentrations. **e**, [³⁵S]GTP- γ S binding to purified G_i, G_o, (30 ng) or G_s (20 ng). **f**, [³⁵S]GTP- γ S binding to G₀ pre-incubated with H₂O₂ (300 μM , 10 min), and then quenched by 10 mM NAC.

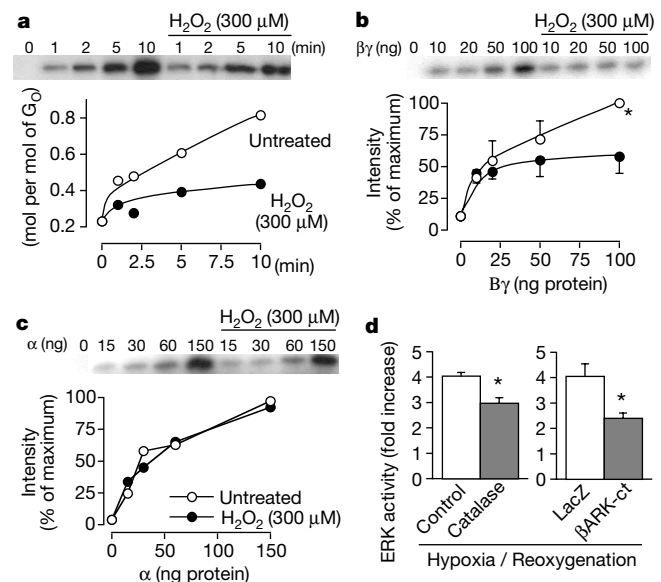


Figure 4 Mechanism and physiological function of H₂O₂-induced ERK activation. **a**, Heterotrimeric G₀ proteins pre-incubated with H₂O₂ (300 μM) for 15 min, were subjected to ADP-ribosylation assay. **b**, **c**, Purified G α_0 in GDP form or $\beta\gamma$ -subunits were pre-incubated with H₂O₂, and ADP-ribosylated by PTX in the presence of the indicated amount of G $\beta\gamma$ (**b**) or G α_0 (**c**). **d**, The cardiomyocytes infected by β ARK-ct or LacZ expressing viruses were exposed to hypoxia for 8 h, and then reoxygenated for 10 min. After the treatment ERK activities were determined. Catalase (25 U ml⁻¹) was included throughout the hypoxia/reoxygenation procedure. Significance from untreated cells (control or LacZ): asterisk, $P < 0.05$.

inhibitors, wortmannin (Fig. 2c) and LY294002 (data not shown). The expression of β ARK-ct did not completely suppress the H_2O_2 -induced activation of ERK and Akt (Figs 1a and 2c). This indicates that a $G\beta\gamma$ -insensitive pathway may also be involved in the H_2O_2 -induced ERK and Akt activation. This idea is supported by reports that other proteins (such as focal adhesion kinase¹⁷ or platelet derived growth factor^{2,4}) may participate in the oxidative stress-induced activation of ERK and Akt.

As β ARK-ct functions upstream of PI(3)K, and PI(3)K is activated by $G\beta\gamma$, we examined whether H_2O_2 directly activates G protein and increases the formation of the GTP-bound form of $G\alpha$ without receptor activation. H_2O_2 increased [³⁵S]GTP- γ S binding in the crude membrane of neonatal myocytes (Fig. 3a, b). This increase was completely inhibited by treatment with an antioxidant, *N*-acetyl-L-cysteine (Fig. 3b). To examine whether heterotrimeric G proteins in the cardiomyocytes are activated by H_2O_2 , we analysed the membranes in a trypsin sensitivity assay. This assay uses the differential digestion pattern of $G\alpha$ subunits between GDP-bound (inactive) and GTP- γ S-bound (active) conformations. Digestion of α_o and α_i , but not α_s and α_q , was limited by the addition of H_2O_2 and GTP- γ S. This resistance indicates that G proteins of the G_i family are mainly activated by H_2O_2 .

To show directly the activation of G_i and G_o by H_2O_2 , we treated purified G proteins with H_2O_2 . H_2O_2 increases GTP- γ S binding to purified G_o in a dose-dependent manner (Fig. 3d). Although GTP- γ S binding to purified G protein was increased by H_2O_2 treatment after a time lag (Fig. 3e), pretreatment with H_2O_2 eliminated this time lag (Fig. 3f). The H_2O_2 -mediated increase of GTP- γ S binding was observed in G_i and G_o proteins but not in G_s (Fig. 3e). We then treated G_o with H_2O_2 , and added *N*-acetyl-L-cysteine in an attempt to cease the action of H_2O_2 . The GTP- γ S binding to pretreated G_o was increased to a magnitude similar to that without the addition of *N*-acetyl-L-cysteine (Fig. 3f), indicating that the modification of G_o by H_2O_2 is irreversible. The H_2O_2 -induced activation of G_o was not affected by the pretreatment of G_o with PTX and NAD (see Supplementary Information). This is consistent with the inability of PTX to affect H_2O_2 -induced activation of ERK in cardiomyocytes, and indicates that H_2O_2 modifies the amino acid in G_o different from that modified by PTX.

To investigate the activation mechanism of G_o by H_2O_2 we analysed G_o treated with H_2O_2 in a PTX-catalysed ADP-ribosylation assay. H_2O_2 markedly decreased PTX-catalysed ADP-ribosylation (Fig. 4a). As ADP-ribosylation catalysed by PTX was measured in the presence of GDP but not GTP, and PTX ADP-ribosylates only the heterotrimeric form of G_i and G_o , this result indicates that H_2O_2 -modified G_o dissociates into $G\alpha$ and $G\beta\gamma$. To determine which subunit is modified by H_2O_2 we first treated $G_o\alpha$ or $G\beta\gamma$ with H_2O_2 and then added non-treated $G\beta\gamma$ or $G\alpha_o$ to measure the ADP-ribosylation catalysed by PTX. Pretreatment of $G\alpha_o$ but not $G\beta\gamma$ with H_2O_2 significantly decreased the ADP-ribosylation (Fig. 4b and c). These data indicate the α -subunits of G_i and G_o are the actual target proteins modified by H_2O_2 . H_2O_2 treatment of $G\alpha_o$ increases GTP- γ S in a time- and dose-dependent manner (see Supplementary Information). These results indicate that modification of $G\alpha_o$ by H_2O_2 may not only increase dissociation into $G\alpha$ and $G\beta\gamma$ but may also accelerate GDP release from $G\alpha_o$. We also found that H_2O_2 treatment increases GTP- γ S binding to recombinant rat $G\alpha_{i2}$ (data not shown). It has been reported that hypoxia/reoxygenation generates H_2O_2 , and activates ERK in a variety of cells⁴. We found that, in cardiomyocytes, hypoxia/reoxygenation increased ERK activity by a mechanism that was sensitive to catalase (Fig. 4d). The expression of β ARK-ct significantly decreased hypoxia/reoxygenation-induced activation of ERK. These results indicate that hypoxia/reoxygenation may generate H_2O_2 and may activate ERK through the $G\beta\gamma$ -dependent pathway.

Many receptors for neuropeptides and hormones activate ERK through both α - and $\beta\gamma$ -subunits of G_i ^{18–20}. We have shown here

that ROS directly activate G_i and G_o without receptor activation. The increase in the GTP-bound form of $G\alpha_i$ caused by ROS leads to liberation of $G\beta\gamma$, possibly activating downstream signal transduction. However, it remains to be determined whether direct activation of G_i by H_2O_2 leads to ERK activation in all cell types. Our results show that the $\beta\gamma$ -subunit liberated from G_i and G_o by oxidative stress activates PI(3)K, which in turn leads to activation of Akt and ERK. As Akt and ERK have been suggested to work together as a survival factor^{17,21}, the intracellular signal transduction induced by the $\beta\gamma$ -subunit of G_i and G_o may protect the cells against oxidative stress. Intracellular redox state changes in various conditions such as ischaemia, stroke, cancer and chronic lung disease, and oxygen radicals are consistently generated by mitochondria and as byproducts of enzyme reactions²². H_2O_2 -induced direct activation of G_i and G_o may therefore be the cellular mechanism that protects the cells against these pathological states. Our report shows that ROS directly activate G_i and G_o , and provides new insight into intracellular signalling by redox regulation. □

Methods

Adenoviruses

We produced recombinant adenovirus for β ARK-ct by the COS-TPC method as described^{11,12}. We obtained recombinant adenovirus for LacZ from the Riken DNA Bank (Tsukuba, Japan). We produced recombinant adenovirus for phosphatase as described²³.

Cell culture and adenoviral infection

We isolated and cultured the ventricles of 1-day-old Sprague–Dawley rat pups as described²⁴. The neonatal cardiomyocytes were plated on 1% gelatin-coated plates at a density of 10^5 cells per cm^2 . Infection with recombinant viruses was performed 24 h after plating at a multiplicity of infection of 100. Forty-eight hours after infection, the cells were exposed to H_2O_2 for the indicated time and collected by lysis buffer containing 10 mM Tris (pH 7.4), 150 mM NaCl, 2 mM EGTA, 2 mM dithiothreitol, 1 mM Na_3VO_4 , 1 mM Phenylmethylsulphonyl fluoride, 10 $\mu g\ ml^{-1}$ leupeptin and 10 $\mu g\ ml^{-1}$ aprotinin. After being centrifuged at 25,000g at 4 °C for 15 min, the supernatants were used for kinase assay and western blotting.

Western blotting

The protein samples (60 μg each) were subjected to 12% SDS–PAGE and electrotransferred onto nitrocellulose membrane. The membrane was incubated with an anti-phospho-specific antibody of JNK (Santa Cruz Biotechnology), p38 MAP kinase (Santa Cruz) or Akt (New England Biolaboratory), and subsequently incubated with horseradish peroxidase-conjugated secondary antibodies. We visualized the reactive bands by the enhanced chemiluminescence method. The total amount of JNK, p38 MAP kinase and Akt were also obtained by rabbit polyclonal antibody (Santa Cruz Biotechnology). We scanned western blots and determined the density of each with NIH image software.

Kinase assay

We determined ERK activity by using BIOTRAK p42/p44 MAP kinase enzyme assay system (Amersham LIFESCIENCE). For the Src kinase assay, we immunoprecipitated cell lysates (200 μg of protein) with anti-Src monoclonal antibody (Upstate Biotechnology); we determined Src activity using Src kinase assay kit (Upstate Biotechnology).

GTP- γ S binding

Neonatal cardiac membranes were prepared as described²⁵. Cell membranes (20–30 μg of protein per tube) were incubated for 20 min at 30 °C in buffer containing 10 mM $MgCl_2$, 2 mM EDTA, 300 mM NaCl, 100 mM Tris-HCl (pH 7.5), 10 μM GDP and 300 nM [³⁵S]GTP- γ S. We collected the cell membrane on a glass fibre filter (Whatman GF/C). For purified G proteins and $G\alpha_o$, 20–30 ng of proteins were incubated in buffer (25 mM HEPES, pH 7.5, 1 mM EDTA, 20 mM 2-mercaptoethanol, 25 mM $MgCl_2$, 100 mM NaCl, 0.7% CHAPS, 10 μM GDP, and 0.5 μM [³⁵S]GTP- γ S), and collected on a nitrocellulose filter (Millipore).

Trypsin sensitivity assay

We carried out trypsin digestion as described²⁶. The membranes (50 μg protein per tube) solubilized by 1% CHAPS in 50 mM HEPES (pH 7.4) were incubated in buffer containing 25 mM HEPES (pH 7.5), 1 mM EDTA, 20 mM 2-mercaptoethanol, 25 mM $MgCl_2$, 100 mM NaCl, 0.7% CHAPS and 10 μM GDP with or without 50 μM GTP- γ S in the absence or presence of 1 mM H_2O_2 for 20 min at 30 °C. We then treated the membranes with 100 $\mu g\ ml^{-1}$ *N*-tosyl-L-phenylamine chloromethyl ketone (TPCK)-trypsin (1:25 ratio of trypsin to total protein) for 15 min at room temperature. The α -subunits of G_i , G_o , G_s and G_q were detected by western blotting.

PTX-catalysed [³²P]ADP-ribosylation

ADP-ribosylation catalysed by PTX was performed as described²⁷. Purified bovine G_o was

first incubated with or without H_2O_2 (300 μM) for 10 min, and *N*-acetyl-L-cysteine was added to cease the H_2O_2 action. Pretreated G_0 was ADP-ribosylated by PTX (10 $\mu\text{g ml}^{-1}$) and [^{32}P]NAD (0.5 μM). We visualized the ADP-ribosylated $\text{G}\alpha_0$ by autoradiography and analysed it by the Fuji BAS system. To determine the activation mechanism of G_0 by H_2O_2 , the individual subunits $\text{G}\alpha_0$ or $\text{G}\beta\gamma$ was first treated with H_2O_2 (300 μM), and then various amounts of $\text{G}\beta\gamma$ or $\text{G}\alpha_0$ were added to measure ADP-ribosylation catalysed by PTX.

Hypoxia/reoxygenation

We performed hypoxia/reoxygenation as described²⁸. The chamber O_2 tension was found to be less than 5 torr during the hypoxic treatment as determined by an oxygen probe (POE-10N, Intermedical, Tokyo, Japan). Hypoxia was maintained for 8 h in the presence or absence of catalase (25 U ml^{-1}). After being reoxygenated for 10 min, ERK activity was measured. Catalase (25 U ml^{-1}) was also added to reoxygenation period.

Statistical analysis

All data were expressed as means $\pm$ s.e.m. We performed three or four independent experiments in duplicate for each condition. Statistical significance was evaluated with Student's *t*-test, and considered to be significant when the *P* value was less than 0.05.

Received 23 March; accepted 29 August 2000.

- Lander, H. M., Ogiste, J. S., Teng, K. K. & Novogrodsky, A. p21^{ras} as a common signaling target of reactive free radicals and cellular redox stress. *J. Biol. Chem.* **270**, 21195–21198 (1995).
- Sundaresan, M., Yu, Z.-X., Ferrans, V. J., Irani, K. & Finkel, T. Requirement for generation of H_2O_2 for platelet-derived growth factor signal transduction. *Science* **270**, 296–299 (1995).
- Yin, Y. *et al.* Involvement of p85 in p53-dependent apoptotic response to oxidative stress. *Nature* **391**, 707–710 (1998).
- Aikawa, R. *et al.* Oxidative stress activates extracellular signal-regulated kinases through Src and Ras in cultured cardiac myocytes of neonatal rats. *J. Clin. Invest.* **100**, 1813–1821 (1997).
- Granger, N. & Korthuis, R. J. Physiologic mechanisms of post-ischemic tissue injury. *Annu. Rev. Physiol.* **57**, 311–332 (1995).
- Downey, J. M. Free radicals and their involvement during long-term myocardial ischemia and reperfusion. *Annu. Rev. Physiol.* **52**, 487–504 (1990).
- Nakamura, K. *et al.* Inhibitory effects of antioxidants on neonatal rat cardiac myocyte hypertrophy induced by tumor necrosis factor- α and angiotensin II. *Circulation* **98**, 794–799 (1998).
- Sugden, P. H. & Clerk, A. Cellular mechanisms of cardiac hypertrophy. *J. Mol. Med.* **76**, 725–746 (1998).
- Luttrell, L. M. *et al.* Regulation of tyrosine kinase cascade by G-protein-coupled receptors. *Curr. Opin. Cell. Biol.* **11**, 177–183 (1999).
- Koch, W. J. *et al.* The binding site for the $\beta\gamma$ subunits of heterotrimeric G proteins on the β -adrenergic receptor kinase. *J. Biol. Chem.* **268**, 8256–8260 (1993).
- Miyake, S. *et al.* Efficient generation of recombinant adenoviruses using adenovirus DNA-terminal protein complex and a cosmid bearing the full-length virus genome. *Proc. Natl Acad. Sci. USA* **93**, 1320–1324 (1996).
- Kurose, H. in *G Protein-Coupled Receptors*. (eds Haga, T. & Berstein, G.) 185–210 (CRC Press LLC, Boca Raton Florida, 1999).
- Koch, W. J. *et al.* Cellular expression of the carboxyl terminus of a G protein-coupled receptor kinase attenuates $\text{G}\beta\gamma$ -mediated signaling. *J. Biol. Chem.* **269**, 6193–6197 (1994).
- Wilkins, J. F. *et al.* Regulation of the kinetics of phosducin phosphorylation in retinal rods. *J. Biol. Chem.* **271**, 19232–19237 (1996).
- Delmas, P. *et al.* $\beta\gamma$ dimers derived from G_0 and G_i proteins contribute different components of adrenergic inhibition of Ca^{2+} channels in rat sympathetic neurons. *J. Physiol.* **518**, 23–26 (1999).
- Coffer, P. J. *et al.* Protein kinase B (c-Akt): a multifunctional mediator of phosphatidylinositol 3-kinase activation. *Biochem. J.* **335**, 1–13 (1998).
- Sonoda, Y. *et al.* FAK is the upstream signal protein of the phosphatidylinositol 3-kinase-Akt survival pathway in hydrogen peroxide-induced apoptosis of a human glioblastoma cell line. *J. Biol. Chem.* **274**, 10566–10570 (1999).
- Howe, L. R. & Marshall, C. J. Lysophosphatidic acid stimulates mitogen-activated protein kinase activation via a G-protein-coupled pathway requiring p21^{ras} and p74^{raf-1}. *J. Biol. Chem.* **268**, 20717–20720 (1993).
- Albas, J. *et al.* Gi-mediated activation of the p21 ras-mitogen-activated protein kinase pathway by alpha 2-adrenergic receptors expressed in fibroblasts. *J. Biol. Chem.* **268**, 19196–19199 (1993).
- Mochizuki, N. *et al.* Activation of the ERK/MAPK pathway by an isoform of rap1 GAP associated with $\text{G}\alpha_1$. *Nature* **400**, 891–894 (1999).
- Bauerle, P. A. & Baltimore, D. NF- κB : Ten years after. *Cell* **87**, 13–20 (1996).
- Semenza, G. L. Perspective on oxygen sensing. *Cell* **98**, 281–284 (1999).
- He, T.-G. *et al.* A simplified system for generating recombinant adenoviruses. *Proc. Natl Acad. Sci. USA* **95**, 2509–2514 (1998).
- Yamashita, N. *et al.* Heat shock-induced manganese superoxide dismutase enhances the tolerance of cardiac myocytes to hypoxia-reoxygenation injury. *J. Mol. Cell. Cardiol.* **29**, 1805–1813 (1997).
- Isogaya, M. *et al.* Identification of a key amino acid of the β_2 -adrenergic receptor for high affinity binding of salmeterol. *Mol. Pharmacol.* **54**, 616–622 (1998).
- Ruiz-Avila, L. *et al.* Coupling of bitter receptor to phosphodiesterase through transduction in taste receptor cells. *Nature* **376**, 80–85 (1995).
- Katada, M., Oinuma, H. & Ui, M. Two guanine nucleotide-binding proteins in rat brain serving as the specific substrate of islet-activating protein, pertussis toxin. *J. Biol. Chem.* **261**, 8182–8191 (1986).
- Kim, K.-O. *et al.* Protection from reoxygenation injury by inhibition of rac1. *J. Clin. Invest.* **101**, 1821–1826 (1998).

Supplementary information is available on Nature's World-Wide Web Site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank T. Katada for providing the purified heterotrimeric proteins (G_i , G_o , G_s) $\text{G}\alpha_0$ and $\text{G}\beta\gamma$. We thank RIKEN DNA Bank for recombinant adenovirus for LacZ. We also thank I. Saito for cosmid vector for the COS-TPC method, and T.-G. He for plasmids of pAdEasy system.

Correspondence and requests for materials should be addressed to H.K. (e-mail: kurose@mol.f.u-tokyo.ac.jp).

Global histone acetylation and deacetylation in yeast

Maria Vogelauer, Jiansheng Wu, Noriyuki Suka & Michael Grunstein

Department of Biological Chemistry, UCLA School of Medicine and the Molecular Biology Institute, Boyer Hall, University of California, Los Angeles, California 90095, USA

Histone acetyltransferases and deacetylases can be targeted to promoters to activate or repress genes. For example, the histone acetyltransferase GCN5 is part of a yeast multiprotein complex that is recruited by the DNA-binding activator protein GCN4 (refs 1–3). The histone deacetylase RPD3 complex is recruited to DNA by the repressor UME6 (refs 4, 5); similar mechanisms exist in other eukaryotes⁶. However, deletion of *RPD3* also increases expression of the *PHO5* gene⁷ that is repressed by nucleosomes^{8,9}, and regulated by GCN5 (ref. 10) but not by UME6. We have determined whether acetylation and deacetylation are promoter specific at *PHO5*, by using antibodies against acetylated lysine residues and chromatin immunoprecipitation to examine the acetylation state of a 4.25-kilobase region surrounding the *PHO5* gene. Here we show that this region is acetylated extensively by ESA1 and GCN5 and deacetylated by HDA1 and RPD3, and that widespread histone modification affects three separate chromosomal regions examined, which total 22 kb. Our data indicate that targeted modification occurs in a background of global acetylation and deacetylation that not only reduces basal transcription, but also allows a rapid return to the initial state of acetylation when targeting is removed.

To determine where histone acetylation occurs in the *Saccharomyces cerevisiae* chromosome 2 domain that includes *PHO5*, we used polymerase chain reaction (PCR) with chromatin immunoprecipitation (Chr-IP)^{11–14}. Using this approach, chromatin from formaldehyde-treated cells is fragmented extensively by sonication, immunoprecipitated by antibodies raised against specific sites of acetylation in histones and amplified using multiplex PCR¹⁵. We designed PCR fragments ranging from 137 base pairs (bp) to 374 bp to analyse a 4.25-kb region encompassing *PHO5* (Fig. 1). We normalized loading to the PCR product of subtelomeric DNA (TEL), which is 500 bp from the end of the chromosome, as it was shown that histone deacetylase (HDAC) mutations do not affect acetylation at this region⁵ (data not shown). Increased PCR amplification of a sequence reflects the increased acetylation at that chromosomal site. We found that in the wild-type strain under repressive conditions (high phosphate medium) H3 lysine 9 (K9; Fig. 1a) and other H3 sites that we examined (K18 and K27; data not shown) are preferentially acetylated in a similar manner at the *PHO5* promoter. H4 K12 (Fig. 1b) and other H4 sites that we examined (K5, K8 and K16; data not shown) are also more acetylated over the promoter region, but to a lesser extent than H3 sites.

As acetylation can be targeted to the *HIS3* promoter¹⁶, the hypoacetylation of the region surrounding the *PHO5* promoter

may be passive, which would reflect a general lack of histone acetylation, or it may be mediated by HDACs. Yeast contains a family of five related histone deacetylase genes: *HDA1*, *RPD3*, *HOS1*, *HOS2* and *HOS3* (ref. 17). To determine whether any of these are involved in deacetylating the chromosomal domain that includes *PHO5*, we performed Chr-IP using wild-type strains and isogenic strains in which each of the deacetylase genes was disrupted. We found that deletion of *HOS1*, *HOS2* or *HOS3* has no obvious effect on histone H3 or H4 acetylation in this region (data not shown). However, *hda1Δ* and *rp3Δ* result in increased acetylation of H3 K9 (Fig. 1a) and H4 K12 (Fig. 1b), which extends over the chromatin domain that includes not only *PHO5* but also the two adjacent genes (open reading frame *YBR094* and *PHO3*).

The widespread acetylation seen when HDACs are deleted suggests that HATs are also acetylating these domains. One such histone acetyltransferase (HAT) may be GCN5, because the constitutive expression of *PHO5* in a *pho80Δ* mutant is abolished by deletion of *GCN5* (ref. 10). Just as *HDA1* deacetylates histone H3 but not H4 *in vivo* (J.W. and M.G., unpublished data), *GCN5* acetylates histone H3 but not H4 *in vitro*¹⁸ and *in vivo* (J.W. and M.G., unpublished data). To determine whether *GCN5* also affects nucleosomes

ubiquitously, we investigated whether deletion of *GCN5* prevents the increased H3 acetylation seen when *HDA1* is disrupted. We found that the acetylation of H3 K9 (Fig. 1a), H3 K18 and H3 K27 (data not shown) over the domain surrounding *PHO5* is abolished by the deletion *hda1Δgcn5Δ*. Acetylation of all four H4 sites is unaffected by either *hda1Δ* or *hda1Δgcn5Δ* (data not shown). Therefore, *GCN5* is responsible for histone H3 acetylation but not H4 acetylation over the 4.25-kb region that includes *PHO5*. *GCN5* is also responsible for increased H3 K9, H3 K18 and H3 K27 acetylation seen in this region when *RPD3* is deleted in a *rp3Δgcn5Δ* strain (data not shown).

ESA1 acetylates preferentially histones H4 and H2A *in vitro*¹⁹. Therefore, using an *esa1^{ts}* (temperature-sensitive) mutant, we investigated whether *ESA1* is responsible for acetylating the sites of histone H4 and H2A that are deacetylated by *RPD3* *in vivo*. We performed Chr-IP on wild-type, *rp3Δ*, *esa1^{ts}* and *rp3Δesa1^{ts}* strains after an 8-h shift to the non-permissive temperature (37 °C). We found that H4 K12 acetylation that would ordinarily be increased when *RPD3* is disrupted is prevented by the *esa1^{ts}* and *rp3Δesa1^{ts}* mutations (Fig. 1b). The increased temperature did not affect acetylation in the wild-type or *rp3Δ* strains (data not

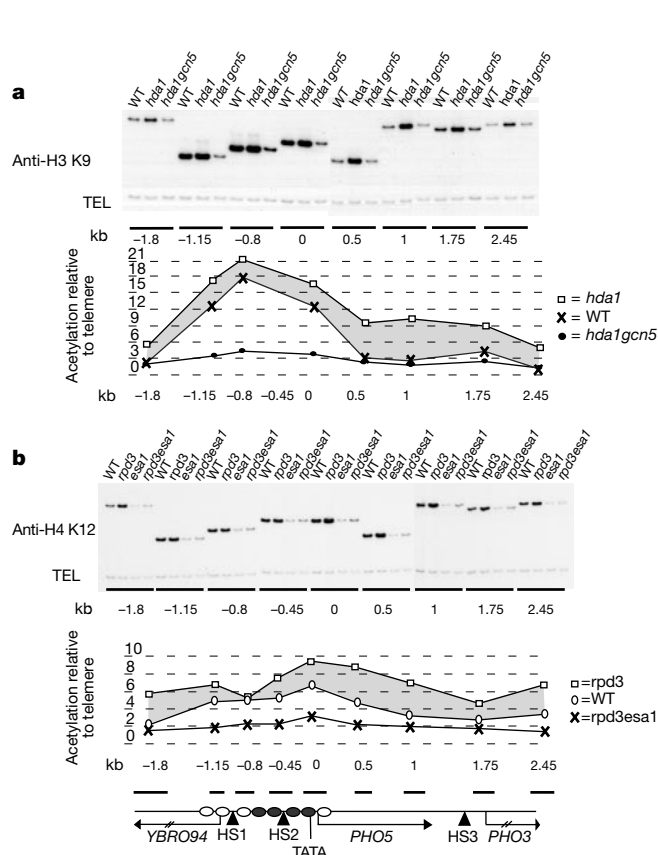


Figure 1 Global acetylation and deacetylation of a chromosomal region containing *PHO5*. **a**, Chr-IP using an antibody against acetylated H3 K9 (anti-H3 K9) and comparison between wild type (WT), *hda1* and *hda1gcn5*. The graph shows relative intensity of the individual fragments after normalization to the telomere specific band (TEL). Increase in acetylation in the *hda1* strain relative to the wild type (shaded) is shown. **b**, PCR analysis of DNA immunoprecipitated by an antibody against acetylated H4 K12 (anti-H4 K12) and comparison between wild type, *rp3Δ*, *esa1* and *rp3Δesa1*. The graph shows relative intensity of the individual fragments of WT, *rp3Δ* and *rp3Δesa1* after normalization to the TEL band. Increase in acetylation in the *rp3Δ* strain relative to the wild type (shaded), and position of the PCR fragments relative to the genomic region (bottom) are shown. Genes and the direction of their transcription (arrows) are shown. HS1, HS2 and HS3 are three sites hypersensitive to DNase I (ref. 28). Positioned nucleosomes (filled ellipses) are shown.

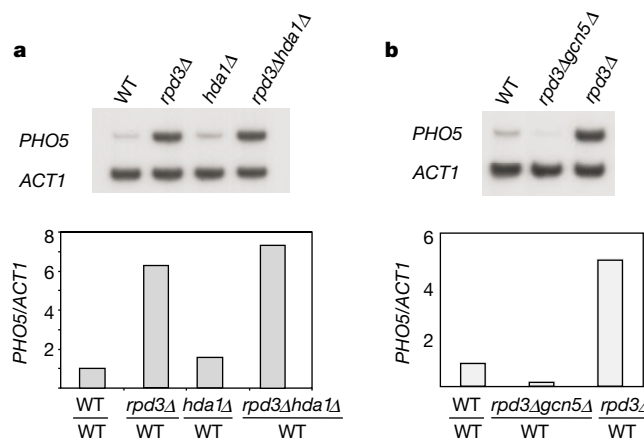


Figure 2 *RPD3* is dominant to *HDA1* in its repression of *PHO5*. **a**, Polyadenylated RNA from wild-type (WT), *rp3Δ*, *hda1Δ* and *rp3Δhda1Δ* strains grown in repressive high phosphate medium. The analysis was carried out by RT-PCR using *PHO5*- and *ACT1*-specific primer pairs. The graph shows the amount of *PHO5* mRNA compared with the wild-type level. **b**, Polyadenylated RNA from wild-type, *rp3Δgcn5Δ* and *rp3Δ* strains. Analysis was performed as in **a**.

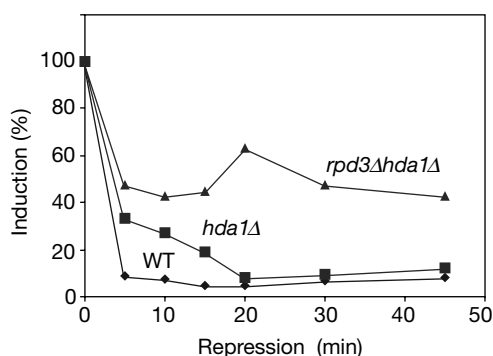


Figure 3 *HDA1* disruption slows return to repression of *PHO5* from activating conditions. Wild-type (WT), *hda1Δ* and *rp3Δhda1Δ* cells were grown overnight in low phosphate activating media. Polyadenylated RNA was prepared from a 5-ml aliquot, and *PHO5* mRNA was analysed by RT-PCR and normalized to *ACT1* mRNA (0 min); the remaining cells were washed twice with water and resuspended in high phosphate repressive media. Aliquots were taken at 5, 10, 15, 20, 30 and 45 min and analysed as above. The percentage of *PHO5* mRNA when compared with the induced level (100%) is shown.

shown). Similarly, we also showed that *ESA1* is responsible for the acetylation of sites K5 and K8 of histone H4, and site K7 of histone H2A (data not shown). Other HATs in yeast include ELP3, which may be associated with the transcription elongator complex²⁰; SAS2, which has a role in silencing heterochromatin^{21,22}; and HAT1, which has been implicated in nucleosome assembly²³. In experiments similar to those above, however, we did not observe significant differences in histone acetylation in the absence of any of these HATs (data not shown). We conclude that *ESA1* and *GCN5* acetylate histone sites that are deacetylated by *RPD3* and *HDA1* throughout the 4.25-kb region that includes *PHO5*.

We next determined the extent to which deacetylation and acetylation affects repression and derepression of *PHO5*. *PHO5* messenger RNA levels were assayed under conditions of repression in strains whose *RPD3* or *HDA1* genes were disrupted. We measured RNA levels by PCR after reverse transcription (RT-PCR) of RNA and quantified the data using phosphorimaging (Fig. 2a). We found that *PHO5* transcription is activated 6.3-fold more by *rpd3Δ* and 1.5-fold more by *hda1Δ*. We observed a 7.3-fold increase in *PHO5* mRNA levels in the *hda1Δ rpd3Δ* strain as compared with wild type. We conclude that *RPD3*-mediated deacetylation of histones H4 and H3 is the main deacetylase function in *PHO5* repression. Moreover, the increased transcription caused by the loss of *RPD3* is abolished completely if *GCN5* is also deleted in an *rpd3Δgcn5Δ* strain (Fig. 2b). Therefore, *GCN5*-mediated acetylation of histones H3 and H2B strongly affects derepression of *PHO5*.

As deletion of *HDA1* has a minor function in repressing *PHO5*, we investigated whether *HDA1* affects the kinetics by which *PHO5* repression occurs. We grew wild-type, *hda1Δ* and *hda1Δ rpd3Δ* strains in low-phosphate-inducing conditions and then transferred them to repressive, high phosphate medium for varying lengths of time. The level of *PHO5* mRNA was measured by RT-PCR as described above. The wild-type strain achieved almost full repres-

sion of *PHO5* after 5 min or less in high phosphate medium (Fig. 3). In contrast, the *hda1Δ* mutant achieved almost full repression at 20 min. *PHO5* is not fully repressed in the *hda1Δ rpd3Δ* mutant strain even after 45 min in high phosphate. This is probably due to the dominant role of *RPD3*. We conclude that increased acetylation caused by *HDA1* disruption prevents the rapid return to a repressive chromosomal state at *PHO5*.

Our above experiments (Fig. 1) analysed *PHO5* in high phosphate repressive conditions. The presence of a peak of acetylation at the repressed *PHO5* promoter indicates that acetylation targeted to the promoter by unknown factors may poise the promoter for subsequent gene activity. Notably, transcription itself does not increase acetylation of H4 at either *GALI* (ref. 5) or *INO1* (N.S. and M.G., unpublished data). Even induction of *PHO5* in low phosphate medium does not increase the acetylation states of H3K9 or H4K12 over the promoter. Moreover, histone acetylase (*GCN5*) and deacetylase (*RPD3* and *HDA1*) disruptions affect acetylation of the entire gene in a similar manner to that seen for the repressed *PHO5* gene (data not shown). Therefore, the acetylation status of *PHO5* and its surrounding region is unlikely to result indirectly from changes in transcription caused when acetylases or deacetylases are disrupted; they are probably a direct result of the disruption in enzymatic activities.

We determined whether other chromosomal regions are acetylated and deacetylated globally by analysing two further regions on chromosome 10, a 10-kb region encompassing the *INO1* gene (Fig. 4a) and a 7.2-kb region chosen randomly (from 145,700 to 153,700 bp, starting at the left end of chromosome 10) (Fig. 4b). We found that disruption of *RPD3* leads to a peak of increased acetylation of H3 K14 at the *INO1* promoter, which is similar to previous data^{4,5}. However, acetylation is also increased throughout the domain that includes *INO1* (shaded, Fig. 4a). The analysis of the 7.2-kb region shows a clear increase in acetylation (shaded, Fig. 4b) over the entire region on disruption of *RPD3*. Therefore, both deacetylation and acetylation are global in these additional chromosomal domains.

The current view of histone acetylation by *GCN5* and deacetylation by *RPD3* involves the targeting of these enzymes to selected promoter elements by DNA-binding proteins⁶. Although some promoters are strongly acetylated or deacetylated in a promoter-specific manner, however, the surrounding regions containing these sites are also acetylated and deacetylated. This global modification may have two important functions. First, the balance of acetylation/deacetylation prevents full acetylation, thereby creating a default underacetylated state. This may decrease basal transcription of many genes by reducing both initiation and elongation of transcription. In fact, a DNA microarray analysis of an *rpd3* mutant showed that 6% of the genome is derepressed by a factor of two or more. This does not include many genes whose basal expression may be increased less noticeably or genes that are already active²⁴. A second function is indicated by the difficulty in returning to the repressed state after *PHO5* induction in cells whose deacetylases are deleted. The rapid turnover of acetyl groups at most nucleosomes may allow chromatin to revert to the initial default acetylation state when targeting is removed. This would prevent the irreversible hyperacetylation or deacetylation of promoters and other regions. □

Methods

mRNA levels

We grew all cells in YEPD or low phosphate-YEPD²⁵ medium before collection. We prepared total RNA by hot-phenol extraction as described²⁶. We purified poly(A)⁺ RNA from 400 μg total RNA using the polyAtract System IV (Promega). We carried out RT-PCR in 10 μl 1 × PCR (Gibco BRL), 3 mM MgCl₂, 1 mM dNTPs, 1 U μl⁻¹ Rnasin (Promega), 0.1 μg μl⁻¹ 9-mer, 0.1 μl poly(A) RNA and 10 U μl⁻¹ M-MuLV reverse transcriptase (New England Biolabs) at 23 °C for 10 min, 42 °C for 30 min and 95 °C for 15 min. We amplified dilutions of the resulting complementary DNA by PCR to ensure amplification in the linear range.

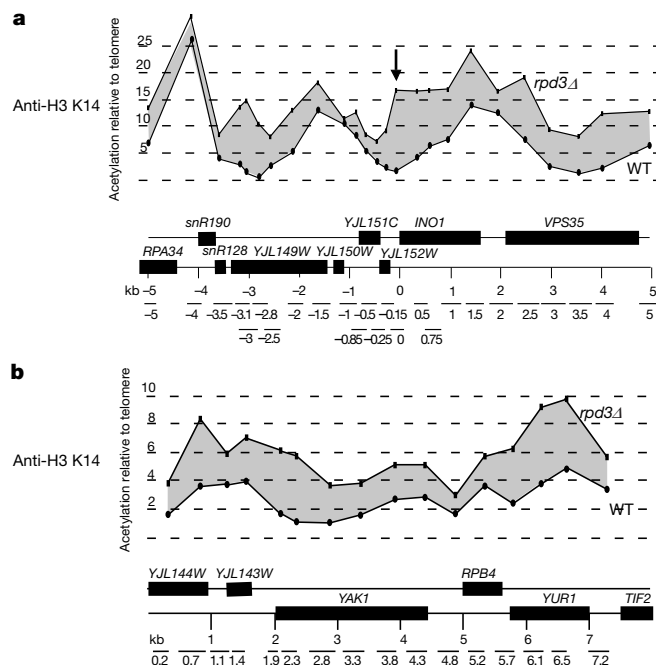


Figure 4 Global deacetylation and acetylation in two regions of chromosome 10. Chr-IP of DNA of wild-type (WT) and *rpd3Δ* strains using an antibody against H3 K14 (anti-H3 K14) was analysed with primers specific for a 10-kb region surrounding *INO1* (a) and an additional 7.2 kb-region at chromosome 10 (b). The graphs show the level of acetylation compared with the telomere-specific band. The shaded area shows the difference between wild type and *rpd3Δ*; location of genes (black boxes) and PCR fragments are shown (bottom).

Chromatin immunoprecipitation

We grew yeast cells to $A_{600} \sim 1$ in YEPD or low phosphate-YEPD²⁵ medium and performed Chr-IP as described elsewhere, using highly specific antibodies raised against individual sites of acetylation and tested in Chr-IP against the relevant histone mutations^{5,27} (N.S. and M.G., unpublished data). We fragmented DNA in chromatin to ~ 500 bp by sonication. We performed immunoprecipitations with the addition of 5 μ l of each antiserum specific for each acetyl-lysine to 50 μ l of whole-cell extract from 8×10^7 cells. We used 1/100 of the immunoprecipitated DNA for analysis by PCR. We performed PCR in the linear range of amplification for each primer set and DNA sample, and added 0.8 μ Ci μ l⁻¹ [α -³²P]dATP (specific activity of 3,000 Ci mmol⁻¹) to each PCR reaction. We quantified the data using PhosphorImager and ImageQuant (Molecular Dynamics).

Strains

Histone acetylase/deacetylase mutants WJY110 ($\Delta hda1$), WJY139 ($\Delta gcn5$), WJY140 ($\Delta rpd3$), WJY141 ($\Delta hda1 \Delta rpd3$), WJY142 ($\Delta hda1 \Delta gcn5$) and WJY143 ($\Delta rpd3 \Delta gcn5$) (J.W. and M.G., unpublished data) are all isogenic to YDS2 (ref. 29). LPY3431 (wild type) and LPY3430 (*esa1*-L327S)¹⁹ were provided by L. Pillus. *RPD3* was deleted in these strains using the *rpd3::LEU2* disruption plasmid pMVL¹⁷ to produce strains MVY006 ($\Delta rpd3$) and MVY007 ($\Delta rpd3 \Delta esa1$ -L327S). We confirmed the disruption by PCR.

Received 29 August; accepted 3 October 2000.

- Georgakopoulos, T. & Thireos, G. Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. *EMBO J.* **11**, 4145–4152 (1992).
- Brownell, J. E. *et al.* Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**, 843–851 (1996).
- Drysdale, C. M. *et al.* The Gcn4p activation domain interacts specifically *in vitro* with RNA polymerase II holoenzyme, TFIID, and the Adap-Gcn5p coactivator complex. *Mol. Cell. Biol.* **18**, 1711–1724 (1998).
- Kadosh, D. & Struhl, K. Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin *in vivo*. *Mol. Cell. Biol.* **18**, 5121–5127 (1998).
- Rundlett, S. E., Carmen, A. A., Suka, N., Turner, B. M. & Grunstein, M. Transcriptional repression by UME6 involves deacetylation of lysine 5 of histone H4 by RPD3. *Nature* **392**, 831–835 (1998).
- Knoepfler, P. S. & Eisenman, R. N. Sin meets NuRD and other tails of repression. *Cell* **99**, 447–450 (1999).
- Vidal, M. & Gaber, R. F. RPD3 encodes a second factor required to achieve maximum positive and negative transcriptional states in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**, 6317–6327 (1991).
- Svaren, J. & Hörz, W. Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. *Trends Biochem. Sci.* **22**, 93–97 (1997).
- Han, M. & Grunstein, M. Nucleosome loss activates yeast downstream promoters *in vivo*. *Cell* **55**, 1137–1145 (1988).
- Gregory, P. D. *et al.* Absence of Gcn5 HAT activity defines a novel state in the opening of chromatin at the PHO5 promoter in yeast. *Mol. Cell* **1**, 495–505 (1998).
- Solomon, M. J. & Varshavsky, A. Formaldehyde-mediated DNA-protein crosslinking: a probe for *in vivo* chromatin structures. *Proc. Natl Acad. Sci. USA* **82**, 6470–6474 (1985).
- Orlando, V., Strutt, H. & Paro, R. Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *Methods* **11**, 205–214 (1997).
- Braunstein, M., Rose, A. B., Holmes, S. G., Allis, C. D. & Broach, J. R. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes Dev.* **7**, 592–604 (1993).
- Dedon, P. C., Soult, J. A., Allis, C. D. & Gorovsky, M. A. A simplified formaldehyde fixation and immunoprecipitation technique for studying protein-DNA interactions. *Anal. Biochem.* **197**, 83–90 (1991).
- Hecht, A., Strahl-Bolsinger, S. & Grunstein, M. Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* **383**, 92–96 (1996).
- Kuo, M. H., Zhou, J., Jambeck, P., Churchill, M. E. & Allis, C. D. Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes *in vivo*. *Genes Dev.* **12**, 627–639 (1998).
- Rundlett, S. E. *et al.* HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proc. Natl Acad. Sci. USA* **93**, 14503–14508 (1996).
- Grant, P. A. *et al.* Expanded lysine acetylation specificity of Gcn5 in native complexes. *J. Biol. Chem.* **274**, 5895–5900 (1999).
- Clarke, A. S., Lowell, J. E., Jacobson, S. J. & Pillus, L. Esa1p is an essential histone acetyltransferase required for cell cycle progression. *Mol. Cell. Biol.* **19**, 2515–2526 (1999).
- Wittschleben, B. O. *et al.* A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Mol. Cell* **4**, 123–128 (1999).
- Reifnyder, C., Lowell, J., Clarke, A. & Pillus, L. Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases. *Nature Genet.* **16**, 109 (1997).
- Ehrenhofer-Murray, A. E., Rivier, D. H. & Rine, J. The role of Sas2, an acetyltransferase homologue of *Saccharomyces cerevisiae*, in silencing and ORC function. *Genetics* **145**, 923–934 (1997).
- Parthun, M. R., Widom, J. & Gottschling, D. E. The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. *Cell* **87**, 85–94 (1996).
- Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
- Han, M., Kim, U. J., Kayne, P. & Grunstein, M. Depletion of histone H4 and nucleosomes activates the PHO5 gene in *Saccharomyces cerevisiae*. *EMBO J.* **7**, 2221–2228 (1988).
- Schmitt, M. E., Brown, T. A. & Trumpower, B. L. A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **18**, 3091–3092 (1990).
- Hecht, A. & Grunstein, M. Mapping DNA interaction sites of chromosomal proteins using immunoprecipitation and polymerase chain reaction. *Methods Enzymol.* **304**, 399–414 (1999).

- Almer, A., Rudolph, H., Hinnen, A. & Hörz, W. Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. *EMBO J.* **5**, 2689–2696 (1986).
- Laman, H., Balderes, D. & Shore, D. Disturbance of normal cell cycle progression enhances the establishment of transcriptional silencing in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 3608–3617 (1995).

Acknowledgements

We thank L. Pillus for the *esa1*^{ts} strain. We are grateful to the members of the Grunstein laboratory for critical comments and discussions throughout this work and are thankful to Y. Suka and A. Carmen for antibodies against histone sites of acetylation. M.V. acknowledges pre-doctoral support from the Dottorato di Ricerca in Genetica e Biologia Molecolare—Università degli Studi di Roma “la Sapienza”. This work was supported by a Public Health Service grant from the National Institutes of Health.

Correspondence and requests for materials should be addressed to M.G. (e-mail: mg@mbi.ucla.edu).

erratum

An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells

David Whyatt, Fokke Lindeboom, Alar Karis, Rita Ferreira, Eric Milot, Rudi Hendriks, Marella de Bruijn, An Langeveld, Joost Gribnau, Frank Grosveld & Sjaak Philipsen

Nature **406**, 519–524 (2000).

In Fig. 2d, GATA-1 tg. Btk^{lacZ/+} females should have been numbered 1 to 3; in Fig. 3a, the last sample should have been marked G4ch no. 14; and in Fig. 4b, the last class of animal should have been tg. female. □

correction

Id2 is a retinoblastoma protein target and mediates signalling by Myc oncoproteins

Anna Lasorella, Michela Nosedà, Mercedes Beyna, Yoshifumi Yokota & Antonio Iavarone

Nature **407**, 592–598 (2000).

The name of one author, Yoshifumi Yokota, was inadvertently omitted from the author list but is included above. He is affiliated with the Department of Biochemistry, Fukui Medical University, Fukui, Japan. □

Material advances

Materials testing, grinding beads and Fractalballs.

Axioplan 2 Imaging

Zeiss

www.zeiss.com

Focused on materials testing

The Axioplan 2 Imaging universal microscope is optimized for metallurgy, topography and materials microscopy, and has an eight-position filter turret with an unrestricted field of 25 mm. All functions are controlled using Axio-Vision software, which enables individual users to save 20 settings that can then be automatically activated by a function key. Options include software packages for grain-size analysis in metals, purity assessment, analysis of cast iron and particle-size analysis in powdered materials.

Reader Service No. 100

StereoEyes Wired

Stereographics

www.stereographics.com

Bring the power of 3D to your PC

StereoEyes Wired enables users to view objects stereoscopically on a PC without the need for specialized graphics cards and software drivers. It is compatible with applications running on Windows 98 and can be used for 3D molecular modelling analysis, GIS spatial analysis and 3D CAD. Other potential applications include medical imaging support, experimental chemistry and PC-based desktop VR.

Reader Service No. 101

Teflon tubing

M. M. Newman

www.mmnewman.com

Flexible and acid-resistant

M. M. Newman Corporation's extruded TFE, FEP and PFA tubing features annular or spiral



Take the tube: Teflon, good for a 900 °F range.



Against the grain: particle size analysis for powder, grain size for metals, using AxioVision.

convolutions, comes with plain or cuffed ends and can be supplied in custom-cut lengths. It is chemically inert, non-flammable, flexible and acid-resistant. Diameters range from 0.25 to 2 inches, and various wall thicknesses are available. The tubing can operate at continuous temperatures from -450°F to $+450^{\circ}\text{F}$.

Reader Service No. 102

CrossFire 2000

MDL Information Systems

www.mdli.com

Hop to it

This search system provides access to the CrossFire Beilstein and Gmelin databases, enabling searches to be carried out on over 8 million organic chemical compounds and more than 1.4 million inorganic and organometallic compounds. The search engine also allows for 'science hopping' — jumping between CrossFire and other applications, so that, for example, selected text in Microsoft Word can be turned automatically into a text-based search query. Users can also define hyperlinks between CrossFire databases and other systems such as proprietary chemical databases and inventory systems.

Reader Service No. 103

Cell-Alert CA-2000

Strison Wireless Systems

www.strison.com

Wireless-to-Internet transmitting unit

This device transmits data from remote locations via the Internet and sends daily reports by e-mail; alarm messages are received in real time

by e-mail and pager. The transmitting unit has six analogue/digital inputs, two counter inputs and one relay output for connection to sensors and other field equipment. Data acquisition, transmission and logging applications are stored in the on-board memory. The unit, with internal keyboard and display, is mounted in a sealed enclosure for installation in harsh environments, and has a back-up battery. Applications include remote monitoring for liquid and bulk solids levels, electric power loss, remote weather, pumps and generators.

Reader Service No. 104

CT 3000

Edax

www.edax.com

Hand-held alloy scanner

The CT 3000 portable alloy scanner rapidly identifies, sorts, matches and provides concentrations of alloys. A pre-loaded comprehensive standards library permits 'out-of-the-box' analysis of metals. Up to 6,000 assays can be performed, stored and analysed in a three-shift operation. The hand-held unit can be used for process control in manufacturing or for quality control.

Reader Service No. 105

ES-7000

Airfiltronix

www.airfiltronix.com

Mini-containment hood

At 17 inches high, 16 inches wide and 17 inches deep, with a top-mounted blower unit, this hood is designed for use where space is limited.

new on the market

It is suitable for adsorption of organic vapours or trapping of airborne powders, smoke or particulates. The blower unit is made of three sections that are attached with a latching mechanism. Gaskets seal each section to provide security against vapour or particulate leakage. A fan speed control permits continuous adjustment of face velocity or airflow. The front panel is made of soft vinyl with Velcro attachments, although a hard or hinged front panel is also available. The unit is suitable for histology operations, local storage of chemicals used in HPLC applications, or mixing small volumes of materials.

Reader Service No. 106

CV0120

Bohlin Instruments www.bohlin.co.uk

Rheometer for solids testing

This rheometer incorporates a DMTA cell that allows full-length solids and delicate films to be mounted in specially designed test clamps. Software enables the user to pre-programme temperature gradients over the range -150°C to $+550^{\circ}\text{C}$ and generates parameters such as glass transition temperature and complex modulus. The sample cell has a range of cone/plate and parallel plate measuring geometries for polymer melts and thermosets. Mechanical spectrum measurements and creep compliance testing produce data that can be used to interpret average molecular weight and molecular weight distribution.

Reader Service No. 107

MicroPatent

Cambridge Scientific Abstracts www.csa.com

New database of materials patents

Subscribers to CSA's Internet Database Service can obtain this new database free of charge. It contains abstracts of US, European and Patent Co-operation Treaty patents related to materials science. The database is updated monthly and currently holds over 60,000 records going back to 1996.

Reader Service No. 108

Fractalballs

Mattech www.buyREFRACTORY.com

Net a bargain

Suitable for a variety of applications, including structural lightweight ceramic membranes, particulate collectors, automotive exhaust applications and filters, these highly porous, high-strength balls are stable up to approximately $1,900^{\circ}\text{C}$. Fractalballs can be bought on-line in a range of sizes and quantities at www.buyREFRACTORY.com.

Reader Service No. 109

Protector

Labconco Corporation www.labconco.com

Controlled atmosphere glove boxes

Protector glove boxes are designed for procedures with oxygen- or moisture-sensitive materials. A large transfer chamber permits materials to be loaded and unloaded without disturbing the environment. Gas input and exhaust of the glove box and vacuum purge and backfill of the transfer chamber are controlled using a four-valve system. A one-piece molded fibreglass or stainless steel liner has radiused corners and a durable finish that resists chemical corrosion. Pressure gauges on the control panel enable the operator to monitor pressures in the glove box and transfer chambers.

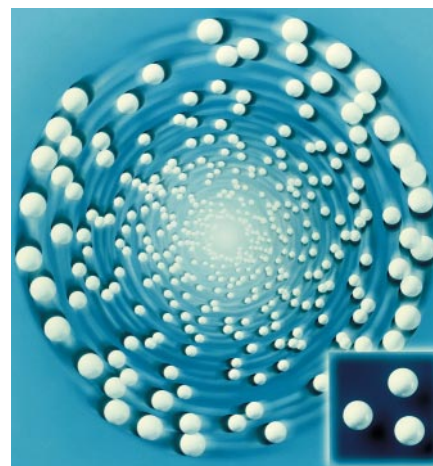
Reader Service No. 110

Zirconia-based beads

Glen Mills www.glenmills.com

Durable grinding beads

Made from high-strength forms of zirconium oxide, these two new grinding beads are claimed to offer mill operators very long bead life, low mill wear, and densities suitable for almost any milling range. The zirconia toughened alumina Type HIG bead has a specific gravity of 4.1+ and is available in sizes ranging from 0.6 mm to 3.3 mm, with other sizes being added to the range. The rare earth stabilized zirconia is formed into tetragonal zirconia polycrystals with a specific gravity of 6.0



Long life: zirconia-based grinding beads

and is currently available in 1.0 mm diameter. The beads are suitable for standard vertical and horizontal mills and the latest basket and high-recycle mills.

Reader Service No. 111

P-10 Automatic Recrystallizer

Design Scientific www.designscientific.com

In search of stability

Polymorphs — multiple crystalline forms of the same molecule — occur in many compounds following crystallization and can affect important properties such as shelf-life, bioavailability, solubility, density, colour and shock sensitivity. They may appear or disappear following production scale-up or changes in impurity levels. The P-10 facilitates rapid identification of the most stable polymorph of a product by forming and dissolving crystals at a rate of 120–600 recrystallizations per day. Users can quickly investigate the formation of solvates or hydrates and achieve maximum purity, thus facilitating production of the most stable polymorph.

Reader Service No. 112

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

TREVIGEN new **PARP**
Poly (ADP-ribose)
Polymerase Reagents

- PARP inhibition assay
- Biotinylated NAD
- Recombinant human PARP
- Anti PARP and PAR

www.trevigen.com
US and Canada: 1 800 873 8443
TEL: 301 216 2800 • FAX: 301 216 2801

READER ENQUIRY NO. 50

**BACs • YACs • PACs
cDNAs**

Research Genetics provides clones and colony membranes for a number of human, mouse and rat libraries as well as a custom screening service for the libraries.

Research Genetics
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

READER ENQUIRY NO. 40
© 2000 Macmillan Magazines Ltd

Look no further
for the best value in
Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!

SIGMA (+44) (0) 1223 839000 • Fax (+44) (0) 1223 839200
email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 33